ACHI 2014

The Seventh International Conference on Advances in Computer-Human Interactions

ISBN: 978-1-61208-325-4

March 23 - 27, 2014

Barcelona, Spain

ACHI 2014 Editors

Leslie Miller, Iowa State University - Ames, USA
Alma Leora Culén , University of Oslo, Norway

ACHI 2014

Foreword

The Seventh International Conference on Advances in Computer-Human Interactions (ACHI 2014), held between March 23-27, 2014 in Barcelona, Spain, continued a series of events targeting traditional and advanced paradigms for computer-human interaction in multi-technology environments. The conference also covered fundamentals on interfaces and models, and highlighted new challenging industrial applications and research topics.

ACHI 2014 was proposed as a result of a paradigm shift in the most recent achievements and future trends in human interactions with increasingly complex systems. Adaptive and knowledge-based user interfaces, universal accessibility, human-robot interaction, agent-driven human computer interaction, and sharable mobile devices are a few of these trends. ACHI 2014 also proposed a suite of specific domain applications, such as gaming, e-learning, social, medicine, teleconferencing and engineering.

We take here the opportunity to warmly thank all the members of the ACHI 2014 Technical Program Committee, as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to ACHI 2014. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the ACHI 2014 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that ACHI 2014 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field of computer-human interaction.

We are convinced that the participants found the event useful and communications very open. We hope that Barcelona, Spain, provided a pleasant environment during the conference and everyone saved some time to enjoy the charm of the city.

ACHI 2014 Chairs:

Leslie Miller, Iowa State University - Ames, USA
Katia Sykara, Carnegie Mellon University, USA
Alma Leora Culén , University of Oslo, Norway
Hermann Kaindl, Vienna University of Technology, Austria

ACHI 2014

COMMITTEE

ACHI Advisory Committee

Leslie Miller, Iowa State University - Ames, USA
Katia Sykara, Carnegie Mellon University, USA
Alma Leora Culén , University of Oslo, Norway
Hermann Kaindl, Vienna University of Technology, Austria

ACHI 2014 Technical Program Committee

Andrea Abate, University of Salerno, Italy
Víctor Acinas Garzón, Instalaciones Inabensa S.A. - Madrid, Spain
Eugenio Aguirre Molina, University of Granada, Spain
Andrea Alessandro Gasparini, University of Oslo, Norway
Florian Alt, University of Stuttgart, Germany
Margherita Antona, Foundation for Research and Technology - Hellas (FORTH) - Heraklion, Greece
Leena Arhippainen, University of Oulu, Finland
Marios Belk, University of Cyprus, Cyprus
Peter Berkelman, University of Hawaii-Manoa, USA
Pierre Borne, Ecole Centrale de Lille / IEEE Fellow, France
Clodis Boscarioli, Western Paraná State University (Unioeste), Brazil
Mohamed Bouri, EPFL, Switzerland
Marius Brade, Technische Universität Dresden, Germany
Lourdes Brasil, University of Brasília at Gama, Brazil
Luis Carriço, University of Lisbon, Portugal
Jessie Y.C. Chen, U.S. Army Research Laboratory - Human Research & Engineering Directorate, USA
Ping Chen, University of Houston-Downtown, USA
Jacek Chmielewski, Poznań University of Economics, Poland
Ryszard S. Choras, University of Technology & Life Sciences - Bydgoszcz, Poland
Thomas Christy, School of Computer Science at Bangor University, UK
Shaowei Chu, University of Tsukuba, Japan
Sharon Lynn Chu, Virginia Tech, USA
José Coelho, University of Lisbon, Portugal
Cesar Alberto Collazos O., University of Cauca - Popayán, Colombia
Andre Constantino, Instituto Federal de São Paulo, Brazil
Sandra Costa, University of Minho, Portugal
Dragos Datcu, Netherlands Defence Academy, The Netherlands
Kerstin Dautenhahn, The University of Hertfordshire, United Kingdom
Lucio Tommaso De Paolis, Salento University - Lecce, Italy
Nancy Diniz, Xi'an Jiaotong-Liverpool University, China
John M. Dolan, Carnegie Mellon University, USA
Rolf Drechsler, DFKI - Bremen, Germany
Carlos Duarte, University of Lisbon, Portugal
Ismael Duque, University of Hertfordshire, UK

Heidi Krömker, Technische Universität Ilmenau, Germany
Satoshi Kurihara, Osaka University, Japan
Andrew Kusiak, The University of Iowa, USA
Jannie Lai, CITRIX, USA
Vidas Lauruska, Siauliai University, Lithuania
Cheongjae Lee, Kyoto University, Japan
Soo-Young Lee, KAIST and RIKEN BSI, Korea
Alma Leora Culén , University of Oslo, Norway
Cynthia Y. Lester, Tuskegee University, USA
Tsai-Yen Li, National Chengchi University - Taipei, Taiwan
Yangmin Li, University of Macau - Taipa, Macao
Chin-Teng (CT) Lin, National Chiao-Tung University, Taiwan
Wei-Kai Liou, National Taiwan Normal University, Taiwan
Jean-Luc Lugrin, Universität Würzburg, Germany
Ameersing Luximon, Hong Kong Polytechnic University, Hong Kong
Francisco Madera, Universidad Autonoma de Yucatan - Merida, Mexico
Jacek Mandziuk, Warsaw University of Technology, Poland
Antonio Maria Rinaldi, Università di Napoli Federico II, Italy
Andrea Mason, University of Wisconsin-Madison, USA
Christoph Mayer, Technische Universität München, Germany
Gerrit Meixner, Heilbronn University, Germany
Harald Milchrahm, Technical University Graz, Austria
Leslie Miller, Iowa State University - Ames, USA
Peyman Moghadam, CSIRO, Australia
Yasser F. O. Mohammad, Assiut University, Egypt
Jaime Muñoz-Arteaga, Universidad Autónoma de Aguascalientes, México
Fazel Naghdy, University of Wollongong, Australia
Toru Nakata, National Institute of Advanced Science and Technology (AIST), Japan
Deok Hee Nam, Wilberforce University, USA
Claudia Nass, Fraunhofer Institute for Experimental Software Engineering, Germany
Andreas C. Nearchou, University of Patras, Greece
Roger Ng, Hong Kong Polytechnic University, Hong Kong
Radoslaw Niewiadomski, University of Genoa, Italy
Gregory O'Hare, University College Dublin (UCD), Ireland
Fakri Othman, Cardiff Metropolitan University (UWIC), UK
Sethuraman (Panch) Panchanathan, Arizona State University - Tempe, USA
Evangelos Papadopoulos, National Technical University of Athens, Greece
Igor Paromtchik, INRIA - Saint Ismier, France
Alexandru Patriciu. McMaster University - Hamilton, Canada
Estelle Perrin, CReSTIC | Reims University – Charleville-Mezieres, France
Igor Petukhov, Blitzkrieg, Russia
Jose Pinto, Universidade do Porto, Portugal
Marina Puyuelo Cazorla, Universitat Politècnica de València (UPV), Spain
Didier Puzenat, Université des Antilles et de la Guyane - Pointe-à-Pitre, Guadeloupe
Rafael Radkowski, Iowa State University, USA
Nitendra Rajput, IBM Research - New Delhi, India
Xiangshi Ren, Kochi University of Technology, Japan
Andreas Riener, University Linz - Linz/Donau, Austria

Copyright Information

Table of Contents

Publicly Displayed Interactive Installations: Where Do They Work Best?

Rune B. Rosseland, G. Snorre Berge and Alma L. Culén
Department of Informatics, Group for Design of Information Systems
University of Oslo
Oslo, Norway
runebro@ifi.uio.no, gsberge@ifi.uio.no, almira@ifi.uio.no

***Abstract*— In this paper, we discuss user experience (UX) with an interactive installation that we have developed in order to study its relation to physical space where the installation is used. The installation utilizes Kinect motion sensor to provide movement based, single or multi user, interactions with graphical and sound interfaces. The installation was tested in various settings, including private interactions in the lab, and public space interactions in a library and a museum. Our findings show that for an open, explorative kind of interactions such as ours, spaces where one is expected to explore, e.g., a Maker Faire or a museum, provide for the longest and most pleasurable interactions with the installation.**

***Keywords*— interactive installations; play; public space; user experience; Kinect.**

I. Introduction

The field of Human Computer Interaction (HCI) can be viewed through the lens of three different paradigms: human factors, classical cognitivist, and the phenomenological, situated paradigm, see Harrison, Tatar and Sengers' work [1]. It is the latter that is of interest for us. It emphasizes a range of more abstract and fuzzy factors that affect HCI, including dynamic use contexts, socially situated action, non-task oriented computing, emotions, etc. [1]. *"It focuses on the experiential quality of interaction, primarily the situated nature of meaning and meaning creation"* [1, p. 1]. This paradigm seeks to produce 'thick', qualitative, subjective, and situated knowledge rather than objective and generalized design rules and models. *"The epistemological stance brought to this site is generally hermeneutic, not analytic, and focuses on developing holistic, reflective understanding while staying open to the possibility of simultaneous, conflicting interpretation"* [1, p. 13].

In particular, user experience, as a field emerging within the phenomenological paradigm, is relatively new and still lacking in theoretical work; see Obrist et al. [2]. Some researchers argue for the use of measurement models and structural models to develop a theoretical understanding of causal aspects of user experience, which can be used to inform design. Others argue for a more holistic approach, where studies of real, situated use are used as the basis for the development of theories. We think this division is artificial. Both approaches are important for an understanding of UX, and should be applied selectively depending on the specific design context. We try to combine the use of theoretical models of user experience with a holistic and open-ended exploration in the wild.

To this end, we have decided to use the increasing interest in public space interactions. Examples of human-technology interactions in public spaces include interactive displays such as large touch based information boards, mobile systems enabling projection in public spaces, tangible interfaces, interactive art and interactive public media. The interaction between a human and the system then becomes public and visible to others who happen to be in the same space. This creates enormous possibilities for interaction and user experience design, yet it is also challenging. One of the challenges we address here is how people feel when exposed to others while trying to interact with the system, see Fig. 1. Many feel silly when making funny gestures in order to interact with or control the system.

Figure 1. Interacting with a nervous robot in the hallway of the school may be intimidating while others observe.

Koppel, Bailly, Muller, and Walter [3] discuss large screens in public spaces in relation to three known major problems: noticing the display, developing motivation for interaction and designing for parallel or collaborative interaction. Their paper looks into configurations of the screen area into flat, concave and hexagonal screens and how these configurations influence users behavior, see Fig. 2. The conclusion they reach is that configurations influence users' behavior: "*Flat created the highest honeypot effect, triggered individuals to position themselves at the extremities of the display, triggered groups to divide and occupy multiple screens, and fostered social learning. Hexagonal allowed strangers to comfortably play on adjacent screens. Concave created the lowest amount of simultaneously interacting people, and caused groups to split into actors and audience",* [3, p. 9].

In this paper, instead of screen configurations in a public space, we consider how the kind of public space influences interactions between people and a system.

Figure 2. Diverse screen configurations lead to different behaviors, [3].

In Section II, we discuss enjoyment, pleasure, play and games. In Section III, we present our methodology, the inspiration for the exhibit, the design concept and the final set up of the exhibit. Users' behavior while interacting with the system in different public spaces is presented in Section IV. Finally, in Section V, we discuss our findings, and, in Section VI, concluding remarks and future work.

II. Enjoyment: Pleasure, Fun and Play

By using enjoyment as an overall category, Blythe and Hassenzahl discuss the semantics of pleasure and fun [4]. Enjoyment can be thought of as an experience fleeting somewhere between distraction and absorption, where, on one end, fun represents distraction, and pleasure represent the absorption side of the scale. In short, fun is described as the counterpart to seriousness. As a distraction, it represents a spontaneous escape from the tasks and worries of everyday life. The self, the hedonic 'be-goals' of UX, does not matter in this short-lived break from reality, but fun still satisfies an important psychological need.

Pleasure is found on the opposite end of the enjoyment scale, taking on the role of absorption. It represents a deeper, longer lasting, more meaningful experience. Here, the connection to people's inner self is made through immersion and devotion to an activity. Elements of challenge, progression, and demand for absolute concentration can be present, and thereby overlaps with Csikszentmihaly's concept of flow; see [5] and [6].

Play is another fuzzy term to corner, as illustrated quite well by Sutton-Smith who has dedicated a whole book to this topic: "*We all play occasionally, and we all know what playing feels like. But when it comes to making theoretical statements about what play is, we fall into silliness. There is little agreement among us, and much ambiguity",* [6, p. 1].

Although the term play represents a myriad of experiences, it has been broadly described as a "*free movement within a more rigid structure*" [7].

Some of the most influential work on play is done by the French sociologist Caillois. He divides play into four forms and two types of play [8]. The four forms of play are competition, chance, simulation and vertigo, and the two types of play are free play and formal play [9].

Playful behaviour is described as an oscillation between exploration and engagement [10]. Playful behaviour starts with exploration, and play occurs when the unfamiliar becomes familiar [11]. When the familiar gets boring, the focus returns to exploration. In this context, the goal of exploring is described as "what can this object do?" and the goal of play is described as "what can I do with this object?".

In relation to our installation, we chose exploration, but in retrospect saw that many of our users would have benefited from having more explicit elements of game. It is possible, as a future work, to change the installation so that users can make explicit choices in terms of more playful or more gameful experiences, according to what gives them a more enjoyable experience.

III. The Interactive Exibit

Our installation focuses on the pleasurable experience for its own sake. It does not solve a problem, nor does it aspire to help people reach meaningful life-goals, though it is designed in such a way that it may be, with minor effort, turned into an exercise installation or a game. Our aim was to design an installation that allowed us to observe and evaluate user enjoyment and behaviour in both public (library and museum) and private (lab) contexts.

A. *The methodology used*

Grounded theory is an inductive research methodology well suited for interpretive research [12, p. 283]. In contrast to the positivist approach of hypothesis development and testing, where the tests are conducted to prove or disprove a predefined hypothesis, grounded theory starts with empirical observations and data, and tries to develop theories from this basis. By grounding our theory development in observations and data gathered from the use of the installation, we attempted to gain some insights into the constitution of enjoyable user experiences and how those experiences are affected by the specific use context and space.

B. *What inspired the installation*

The motion detection was a starting point. An important initial part of the project development was a few months dedicated to exploration of what other similar projects have achieved with installations involving the Kinect sensor or related technologies and concepts. This included scouring the web for videos, tutorials, project description, tools, and examples. A great starting point to look for such resources is the Creative Applications website [13], which is a digital art blog, focusing on *"...projects, tools and platforms relevant to the intersection of art, media and technology*". The most influential pieces for the present work were the *V Motion Project*, based on creating music through motion; and the Firewall installation, based on manipulating the speed and volume of music by touching a stretched sheet connected to sensors. In addition, we were inspired by calligraphic movements, martial arts, and different dance styles.

C. *Physical set up of the installation*

In all locations, the installation was exhibited in the setup as shown diagrammatically in Fig. 3, and in the actual space in Fig. 4. Each of the locations had, at least, an area of four by four meters in front of the Kinect sensors. The installation consisted of:

- A long and narrow table placed by a wall.
- Two Kinect sensors mounted on the table on top of each other.
- One Shake 'n' Sense device [14], fastened to one of the sensors to eliminate interference.

 ISBN: 978-1-61208-325-4

- A wall-mounted screen, either a flat screen TV or a canvas lit by a projector.
- Two amplified speakers placed on the table on each side of the screen.
- Two Mac laptops placed outside of the installation area, one running the audio and the other the visual system.

The final installation consisted of two completely separate systems, one controlling the audio and one controlling the visual display. This separation was made to keep the systems as stateless as possible. By stateless we mean that the systems did not keep or store any information on previous states or actions. This means that the systems reacted only to what a user was presently doing, and was not based on what has been done earlier by this user or any other user. The systems were tuned to work together and appeared for the user as a single installation.

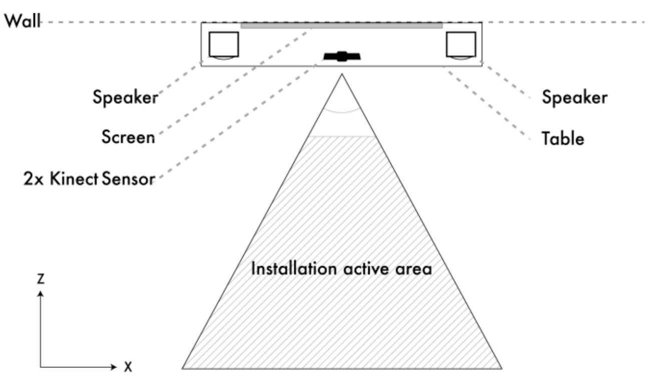


Figure 3. Diagram representing the physical setup of the installation.

Figure 4. Actual setup of the installation in the museum during the Maker Faire, with interaction area marked on the floor.

When a user, or several users, walked into the range of the sensors, the system automatically calibrated them and started tracking their movement and playing the sound track.

The main way a user could start interacting with the installation was by extending an arm away from their body. More specifically, a horizontal hand movement away from the chest would trigger the system and start the calibration processes, engaging both audio and visuals. The distance for this horizontal hand movement was set to 25 cm.

IV. THE SPACE AND THE USER BEHAVIOUR

The first public test of our installation, apart from one exhibit in the lab open to general audience, was at the Science Library at the University of Oslo. The library actively encourages students to develop different kinds of systems and technologies to be used in the library. We were invited to set up our installation in the foyer on the ground floor of the library building for three consecutive days. This provided a good opportunity for us to observe how people reacted to and interacted with our installation in a real public setting; see Fig. 5.

Figure 5. The installation in use in the Science Library at the University of Oslo.

At the library, two researchers were present to observe and take notes. In order to ensure that notes would be comparable, a simple coding system was designed. Seven different pieces of information were recorded for each user engagement: something to identify a person or a group by; start- and end- times; single user or a group; if they actively engaged with the installation or just passively observed; body language; facial expression; and finally, notes. Body language was categorised into: shy, curious, engaged, uninhibited, frustrated, self-conscious, indifferent, joyful, and sceptical. Facial expressions were recorded as emoticons.

The observations were done over a three-day period, in two different locations within the library. The first two days, the installation was placed in the foyer of the library, in the vicinity of the cafe. This was the most trafficked area in the library. The third day, it was moved to a more quiet area. The set up was as follows:

- Duration of observation: 7 hours 49 minutes over three days (2h + 2h 36m + 3h 13m).
- People observed: more than 52 (52 interaction sessions, some of them involving groups).
- Active: 35 (if a group, all participants took an active role)

- Passive: 17 (if at least one member of a group was just observing)
- Age: mainly from 17-40 years old, a few in forties, fifties or sixties.

The granularity of our time-registrations were not fine enough to draw any certain conclusions regarding time spent with the installation, other than that hardly anyone spent more than three minutes. However, when comparing time spent to whether the person was alone or in a group, we saw that people who were in a group spent more than twice as long (1.2 minutes in average) than a person who interacted with the installation alone (0.5 minute in average).

Perhaps in contrast to the usual absence of music in the library, people soon learned that whenever the music started, there were people interacting with the installation. This allowed them to look up whenever the installation was in use, thereby slowly building an understanding of how it worked. This also allowed them to build both curiosity and courage to try the installation for themselves. We saw several examples of people coming up to investigate after having observed others interacting with it for a while. There were also examples of single persons and groups of people who were hanging around in the background, queuing when others were interacting with the installation. As soon as the people using the installation left, they would walk up and give it a try. This worked like a honeypot effect, a positive feedback loop, where use attracted attention and instigated more use. However, the installation was unable to keep people's interest for more than a minute or two, which meant that there would have to be a constant stream of people to keep the installation in continuous use. When the installation was allowed to go into standby mode, people quickly returned their attention to whatever they were otherwise doing.

In terms of level of engagement, those people who explored the installation together with others seemed to get more out of it than those that were alone. They would talk to each other and explore cooperatively, discovering more functionality than those that were alone. There were also several examples of people who had been interacting with the installation earlier came back with friends.

Verbal reactions were usually immediate and short, perhaps also because the observers were hidden, looking just like everyone else, so people were more or less talking either to themselves or to their friends:

"Awesome! Motion sensor, cool!" – Man X

"Shit! Wow!" – Girl A

"Very cool!" – Man Y

"Pretty cool!" – Man Z

There were also more reflective statements:

"It responds to my movement." – Man W

After exploring for a minute, one man, of about 60 years old, exclaimed: *"One could stand here all day, fooling around!"* – Man P.

Many participants have thus explicitly mentioned the word cool. Coolness of technological objects may be an important factor for their consequent acceptance; see [15].

The second public test took place at the Norwegian Science Museum, during the Mini Maker Faire in Oslo; see Fig. 4 and Fig. 6.

Figure 6. Interacting with the system at the Norwegian Science Museum.

Doing the same coding as for the library, we found the following:

- Duration of observation: 42 minutes, one day.
- People observed: 33.
- Active: all
- Age: from about one year old to somewhere in the sixties.

Average interaction time was over 2 minutes.

V. DISCUSSION OF THE GENERAL FINDINGS

In this section, we will discuss how the user experiences appeared to be affected by the different public contexts the installation was tested in.

The audience at the library included mostly students, faculty, and other staff. They were there because they had some business there, either going to or from a lecture or the library, hanging out with friends, studying, eating lunch, and so on.

At the museum, on the other hand, there was a greater mix of people, ranging from toddlers to grandparents, but with an overweight of children in the pre- and primary school ages. They mostly arrived in groups, with family members or friends, and were there to experience, learn, and enjoy themselves.

The physical spaces our installation was exhibited in were quite different. The two locations at the library were exposed and crowded, particularly the first one. This meant that anyone interacting with the installation would draw attention from not only the immediate surroundings, but also from galleries on the floors above. The sound would naturally draw attention from the surroundings, and given the open layout of the building, it was allowed to disperse throughout the building. Furthermore, as we accidentally discovered when the installation was run without sound output, people did not take notice of the installation at all when it did not produce any sound. This may be explained by the term *display blindness*; people have become so used to all kinds of public displays and advertising that they can selectively ignore them. *Interaction blindness* refers to the fact that it is difficult for people to understand whether a given display is interactive. Houben and Weichel [16] have described how display blindness and interaction blindness can be overcome by use of curiosity objects, e.g., objects that are designed to draw attention by sparking interest and curiosity. Although not intended specifically as a curiosity object, the sound certainly worked as one, effectively drawing attention to the installation whenever it was triggered.

The space we were assigned at the museum was partly confined, making it close to impossible for others to observe the installation, or the people interacting with it, from afar. This seemed to give participants a sense of privacy and allowed them to let themselves get more carried away than at the library. Also, our preparation of the installation space with chairs for onlookers to sit on along the sides was very beneficial. It allowed the ones who did not want to try the installation to sit down and relax, but still be able to communicate and take part in the experience with their friends who were interacting with the installation. Several onlookers also eventually got up and tried the installation after having grown curious by watching others.

A very central concern for many people seem to be a reluctance to appear conspicuous or out of the ordinary in public spaces. Breaking social rules and norms is a big deal. As Roto et al. [17] pointed out, *"UX is rooted in a social and cultural context"* .

Most of our interviewees at the prototype evaluations in the lab readily admitted that they would restrain their involvement with the installation in a public setting, if they would be willing to interact with it at all. The most central reason they gave for this was the fear of breaking social rules and norms, and of "behaving like an idiot", as one participant put it. It was their fear of being perceived by others as doing something people do not normally do in public that would keep them from getting too involved. There were also comments to the opposite effect, indicating that breaking social rules and norms can be liberating and empowering. However, the prevailing notion was that social rules and norms would have a dampening effect on people's level of involvement with such an installation in public settings.

This concern seemed particularly evident at the library. The openness of the location and the number of people in the surrounding area seemed to make people self-conscious and vulnerable when they triggered the installation, particularly if they were alone. At the museum, there was clearly more headroom for expansive and expressive behavior. Many of the permanent museum exhibitions are designed for interaction and exploration, and the wide variety of strange projects taking part in the Maker Faire clearly made people less concerned about how their behavior would be perceived by others, as this behavior was expected in this context. Nevertheless, there were examples at both locations of people showing an interest in the installation but being too shy to dare to try it for themselves. But by having the opportunity to watch others interact with it and build an understanding of how it worked, the shyness was often overcome by curiosity, resulting in them engaging with the installation after having observed for a while.

Next, we present the general findings from the evaluation sessions and public tests of our installation. Through the analysis of our evaluation sessions in the private lab context, we located many statements indicating an enjoyable experience. In this section, we want to look at aspects of the participants experiences related to the concepts encompassed by the term 'enjoyment', as discussed in Section 2.

A. Fun

The installation in itself was described by most as 'fun'. Blythe and Hassenzahl defined fun as a short-lived distraction from everyday life [16], coinciding well with the way the word is used in describing the experience by the participants. But, what exactly was fun about the installation? The participants answers points first and foremost to the exploration of the installation and its functionality, then secondly, the immediate responses the installation gave to movement, and the sensory aesthetic experiences they resulted in.

Pleasure was never mentioned directly by the participants, but several interviewed participants described an experience of 'flow' [18] when they were interacting with the exhibit, which can be linked to pleasure [4]. These experiences were described in terms of being 'lost', mesmerized, having a mental break and entering a relaxed 'kind of mode', and the majority of the participants agreed on this being an essential part of their experience. It's worth noting that some of the participants pointed out, both explicitly and implicitly, that this flow-like state disappeared over time as the participants ran out of elements of the installation to explore.

B. Play

Several of the participants described the installation and the experience as playful. Their descriptions indicated that they placed the experience more in line with the definition of free play, rather than formal play (game).

The playfulness the installation facilitated for was deemed as very important, and the participants linked it strongly to the exploration part and the open-endedness of experience, but also to the lack of control. The openness of the installation was described as an advantage, in the way that it encouraged interpretation and exploration. The lack of control was described as not important by one participant, as the point is not to steer something, but to play with the system and get responses from it, which resulted in a 'good feeling'. In relation to the concepts of goals, rules, and competitive elements of play, even the self-proclaimed 'competition-focused' participants acknowledged that those concepts were not the point of this installation.

In the playful behavior, there is an oscillation between exploration and play, where exploration is triggered by boredom [10], [11], [19] and play is triggered by learning or discovery [20]. We found multiple instances of this in the way participants described their explorative behavior, which strongly resembles the process of playful behavior, emphasizing the strong relation between playing and exploring: *"It is just exploring, really. Until you feel you master (the installation) a bit, then it's really exciting and makes you want to continue. You never know if you have explored everything and that's positive, you never reach an end."*

C. Aesthetics

In terms of aesthetics, both the audio and the visuals were described as fascinating, atmospheric, different, beautiful and soothing. The participants thought the combination of the two fit well together and resulted in a coherent expression and created a good ambiance. It was also pointed out in a positive manner that the expression was kept to an abstract nature. That way it became easier to accept the audiovisual expression, in comparison to trying to depict or simulate something concrete.

D. Exploration

As stated earlier, exploration was the activity and experience deemed the most important successful aspect of the installation. Several of the participants expressed bluntly that exploration *is* the installation. The exploration was fuelled by the responses given by the installation and their abstract, mysterious, unknown nature. Or, to put in other words, the immediate responses to movement and actions, combined with lack of explanation, made the participants curious and eager to investigate. Their descriptions also highlighted one of the common characteristics of the human brain, namely the constant search for patterns and connections, which was described as an essential part of the process of exploring.

E. Discovery, learning and understanding

On some aspects of the experience, the participants were quite divided in their opinions. One of these aspects was the lack of explanation, or guidance, in the user interface of the installation. The majority of participants highlighted the absence of explanations as something positive. It was seen as a catalyst for, and a component of, exploration. However, two of the participants found it confusing, frustrating and incomprehensible.

One of the participants, who favored minimal explanations, pointed out that an installation such as ours would not be suitable for people who are not interested in exploring.

The discovering and learning were described as closely related to exploration. For example, one participant described discovery as a direct result of the exploration.

The process of understanding was the challenging part of the installation. When exploration led to discoveries and understanding, the participants had a sense of progress and achievement, giving them motivation to continue to explore. However, the lack of 'new things' to discover and explore eventually led to boredom and loss of interest.

Some users considered the lack of progression and control as negative aspects of the installation.

F. Progression

Progression was an aspect of enjoyable user experience that was originally overlooked by us in the design process, but which surfaced through the evaluation of the prototype as the most important missing aspect of the participants' experience. As mentioned earlier, Blythe and Hassenzahl link the concept of pleasure to the concept of flow, but they also argue that pleasure can in fact be thought of in terms of progression [4]. In retrospect, this actually comes across as self evident, when comparing our findings to the overlapping definitions of flow and pleasure, as a longer lasting, more meaningful and immersed experience devoted to an activity.

The participants wanted more depth to the experience. They wanted more to explore, and gradually increasing variation and difficulty. When they felt they had exhausted their possibilities for exploration, they became bored, and this coincided with the earlier mentioned loss of flow state.

G. Control

The second most sought after aspect was control, and on this topic the participants of the prototype evaluation was close to unanimous. They expressed frustration over not getting the expected responses from the system, and this put limitations on what they could do. It prevented them from being creative and expressing themselves through the installation, both in terms of visual and audio expression, and this was emphasized as important to them. Some acknowledged that they attained a certain degree of control, but it was expressed that the threshold for gaining this control should be much lower in order to make the installation accessible to more people.

The lack of control linked very strongly to the absence of mastery, and on this point the feedback from one of the users was quite direct: *"*[The installation] *lacks possibility for mastery."* And another user on the same topic: *"I don't think I would master it more if I used it for another 20 minutes."*

The only positive feeling described that related to mastery was through exploring and discovering, but even then, this process was described as fumbling.

The lack of control highlights the relation between the second- and third- paradigms of HCI, and between usability and user experience. In our phenomenological approach, the focus was on enjoyable user experiences, and not so much on usability and ease of use on a specific level. Also, the explorative and abstract nature of the installation meant that it was difficult to define specific usability criteria for it.

This is not to say that control was not a focus in our design, but the lack of precision in the tracking data from the Kinects, and our experiential focus led us to design a system that did not need very specific and precise controls. Nevertheless, our findings clearly show that lack of control detracts from the experience. This is in line with the fact that good usability is a prerequisite but not a guarantee for good user experiences. It also underlines the importance of both second- and third- paradigm HCI approaches for the overall user experience; neither approach is complete on its own, but must be combined selectively according to the

 ISBN: 978-1-61208-325-4

specific design context in order to make the user experience as enjoyable as possible.

VI. DISCUSSION OF THE FINDINGS RELATED TO SPACE

We now focus on presenting the space related findings specifically.

In terms of time spent by participants interacting with the installation of the two locations, we see the time spent at the museum was significantly higher. At the library, no one spent more than three minutes with the installation, 41% spent less than one minute and 72% of the observed spent two minutes or less. At the museum, the time spent with the installation is spread much more evenly across the intervals noted: 59% spent two minutes or more interacting, and some people seen outside the time frame of observations were exceeding the intervals noted significantly.

Looking at the distribution of facial expressions observed in the two different contexts (Fig. 7 and Fig. 8), expressions of a positive nature are the predominant ones in both settings, but at the museum as many as 86% were smiling and even though 5% were noted as indifferent, 95% of the observed were deemed positive.

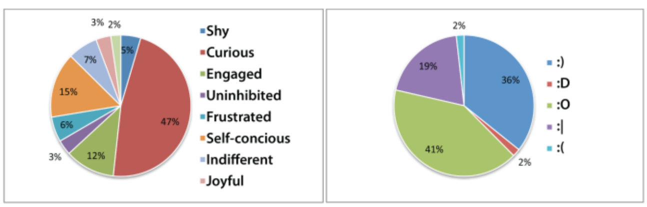


Figure 7. Body language and facial expression distributions at the library.

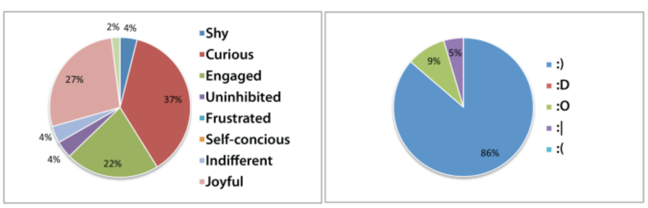


Figure 8. Body language and facial expression distribution at the museum.

Comparing the observations of body language between the contexts, a high degree of curiosity is observed in both settings, with 47% recorded as displaying a body language suggesting curiosity in the library setting, while 37% were recorded at the museum. The most striking difference between the library and the museum contexts was the high percentage of joyfulness (27%) and the low percentage of shyness (4%) of the museum setting, contrasting the low degree of joyfulness (3%) and high degree of self-consciousness (15%) and shyness (5%) (combined 20%) seen at the library. The reason for combining self-consciousness and shyness is that they are very similar traits. Seen in retrospect, separating these terms into two coding categories might have been unnecessary, considering their similarities and the fallibility of observation.

During our prototype evaluations, testers consistently underlined that they would be less likely to interact with the installation on the street or in the shopping center than if they encountered it at a destination like a museum or a gallery. One of the participants expressed this as follows: *"It would be a lot more socially acceptable in a museum to interact with it. I would say my experience would have been much better in a context like that. If the installation were set up in Karl Johan* [note: central shopping street in Oslo] *I wouldn't have stopped to check it out, also because I'm going somewhere".*

Applying this thinking to the library and museum, we see that the foyer of the library is a place which one passes through on the way. The museum is a destination in itself. The library foyer is thus more similar to a 'street' setting, in that the installation is unexpected and disrupts passers-by. At the museum, on the other hand, visitors expect to explore and have an experience. If we also take into account the fact that people spent considerably longer time interacting with the installation at the museum (Fig. 8) than in the library, this might support the notion that the museum / Maker Faire setting was a better-suited context for the intended use of the installation, namely facilitating for spending time exploring. An interesting observation is that the amount of people marked as being fascinated by the exhibit varies significantly between the two observation places. At the library, 41% were recorded as fascinated, in contrast to only 9% in the museum context. This might suggest that fascination was expected in the context of the museum, while unexpected in the context of the library.

If we look at the distribution of the differences observed in the body language (and especially the differences between joyfulness and the combination of self-consciousness and shyness observed in the two contexts), it seems that the library context was perceived as a less comfortable one. We attribute this to the 'deviant' nature of the interactions from what is the norm at the library. Waving their arms in mid-air makes participants of the installation stand out and calls attention to them from people situated nearby who are not aware of what the participants are doing, thus making many participants uncomfortable. This situation was predicted by some participants of the private evaluation sessions. They said it was important for them to make their actions understandable to onlookers, if they were to engage with such an installation in a public space.

These suggested effects of breaking social rules and norms and disruption (or surprise) indicated, especially by the observations in the library, are not necessarily to be considered negative. We may consider the installation, as it worked at the time, better suited for the museum / Maker Faire context, but through alterations to the installation based on feedback from the evaluation sessions, we imagine we could have facilitated a better user experience at the library as well. This could be to make the 'tools' or controls more obvious to help the shy/self-conscious but curious understand the basic workings before they reach the stage of giving up, preventing potential frustration and embarrassment. To help avoid the perception of 'unnatural' behavior to onlookers, we could make the connection between participants and their screen presence much clearer, hopefully making it more obvious to the onlookers what the participants were doing.

VII. CONCLUSION AND FUTURE WORK

We have designed and implemented an audiovisual interactive installation for the purposes of exploring playful and pleasurable interactions in public spaces.

For the purposes of this paper, we explored the differences in user behaviors in two distinct public setting: at the university library and in the science museum, during a Maker Faire.

The results show that the context in which the exhibit is installed and used strongly influences user's behavior. Exploration, discovery and learning in public space need to be supported properly. The museum is the context suitable for such endeavor and participants have used longer time to interact with the exhibit, have shown much less frustration, were less concerned with on-lookers, and showed clear signs of pleasure, e.g., smiling. However, in the museum setting, it was harder to fascinate the users of the exhibit. The visitor's expectations in this regard are high. People in the library are there on other accounts and thus are harder to engage, more shy but easier to solicit a "wow" effect from.

Our evaluation of the installation was qualitative, and it was analyzed through applying grounded theory techniques, such as coding. The analysis gave us valuable insights, also highlighting some weaknesses of the installation in terms of, for example, the lack of control and progression during interaction.

We envision several possibilities for future work based on the installation, such as modifying it into a game. Furthermore, some important concepts from UX theory may be studied through interaction with the installation, such as immediacy of understanding and affordance, and how they are influenced by the fact that several people may interact with the installation simultaneously.

REFERENCES

[1] S. Harrison, D. Tatar, and P. Sengers, "The three paradigms of HCI," in Alt. Chi. Session at the SIGCHI Conference on Human Factors in Computing Systems San Jose, California, USA, 2007, pp. 1–18.

[2] M. Obrist, V. Roto, A. Vermeeren, K. Väänänen-Vainio-Mattila, E. L.-C. Law, and K. Kuutti, "In search of theoretical foundations for UX research and practice," in CHI '12 Extended Abstracts on Human Factors in Computing Systems, New York, NY, USA, 2012, pp. 1979–1984.

[3] M. T. Koppel, G. Bailly, J. Müller, and R. Walter, "Chained displays: configurations of public displays can be used to influence actor-, audience-, and passer-by behavior," in Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems, 2012, pp. 317–326.

[4] M. Blythe and M. Hassenzahl, "The semantics of fun: differentiating enjoyable eeperiences," in Funology, M. A. Blythe, K. Overbeeke, A. F. Monk, and P. C. Wright, Eds. Springer Netherlands, 2005, pp. 91–100.

[5] M. Csikszentmihalyi, Flow: the psychology of optimal experience. New York, N.Y.: HarperPerennial, 1991.

[6] B. Sutton-Smith, The ambiguity of play. Cambridge, Mass.: Harvard University Press, 1997.

[7] K. Salen and E. Zimmerman, Rules of play: game design fundamentals. Cambridge, Mass.: MIT Press, 2003.

[8] R. Caillois, Man, Play, and Games. University of Illinois Press, 2001.

[9] B. Costello and E. Edmonds, "A tool for characterizing the experience of play," in Proceedings of the Sixth Australasian Conference on Interactive Entertainment, New York, NY, USA, 2009, pp. 2:1–2:10.

[10] B. Costello and E. Edmonds, "A study in play, pleasure and interaction design," in Proceedings of the 2007 conference on Designing pleasurable products and interfaces, New York, NY, USA, 2007, pp. 76–91.

[11] J. N. Lieberman, Playfulness: its relationship to imagination and creativity. New York: Academic Press, 1977.

[12] D. J. Lazar, D. J. H. Feng, and D. H. Hochheiser, Research methods in human-computer interaction. John Wiley & Sons, 2010.

[13] "CreativeApplications.Net | Apps That Inspire..." [Online]. Available: http://www.creativeapplications.net/. [Accessed: 30-Jan-2014].

[14] A. Maimone and H. Fuchs, "Reducing interference between multiple structured light depth sensors using motion," in 2012 IEEE Virtual Reality Short Papers and Posters (VRW), 2012, pp. 51–54.

[15] A. L. Culén and A. Gasparini, "Situated techno-cools: factors that contribute to making technology cool and the study case of iPad in education," PsycNology Journal, vol. 10, no. 2, pp. 117–139, 2012.

[16] S. Houben and C. Weichel, "Overcoming interaction blindness through curiosity objects," in CHI '13 Extended Abstracts on Human Factors in Computing Systems, New York, NY, USA, 2013, pp. 1539–1544.

[17] V. Roto, E. Law, A. Vermeeren, and J. Hoonhout, "User experience white paper - bringing clarity to the concept of user experience," presented at the Dagstuhl Seminar on User Experience - 2010, 2011.

[18] "Go With The Flow." [Online]. Available: http://www.wired.com/wired/archive/4.09/czik_pr.html. [Accessed: 30-Jan-2014].

[19] B. M. Costello and E. A. Edmonds, "Directed and emergent play," in Proceedings of the seventh ACM conference on Creativity and cognition, New York, NY, USA, 2009, pp. 107–116.

[20] A. L. Culén, T. Bratteteig, S. Pandey, and S. Srivastava, "The Child-to-Child (C2C) Method: Participatory Design for, with and by Children in a Children's Museum," *IJWI*, vol. 11, no. 2, pp. 92–113, 2013.

Using the Implicit Association Test for Interface-Based Evaluations

Tiago Devezas
Research Center for Assistive Information and Communication Solutions (AICOS)
Fraunhofer Portugal
Porto, Portugal
tiago.devezas@fraunhofer.pt

Bruno Giesteira
Faculty of Fine Arts
University of Porto
Porto, Portugal
bgiesteira@fba.up.pt

***Abstract*— Non-instrumental dimensions, the aspects of a product that go beyond its ability to help achieve goals efficiently, are increasingly important in User Experience (UX) research. These dimensions, which include qualities like aesthetics and symbolism, are mainly assessed by self-reports, research has shown. However, respondents can provide wrong answers, willingly, due to concerns like social desirability and self-presentation, or unwillingly, due to the inability to access their inner states. We explored if one implicit measuring method, the Implicit Association Test (IAT), can be used to complement or replace self-report measures. Participants completed six IATs and explicit measures to determine their attitudes toward products represented by pictures of their interfaces. Two non-instrumental dimensions were assessed: valence and self-identification. Overall, implicit and explicit measures displayed a medium correlation. When comparing the correlations between the IATs for the two assessed dimensions and the corresponding explicit measures, similar strong effects were found. This suggests that the IAT bears further exploration as a complement or alternative to self-report methods.**

Keywords-Implicit Association Test; interface evaluation; aesthetics; User Experience

I. INTRODUCTION

In this paper, we explore if one implicit measuring method, the IAT, is a valid complement or alternative to explicit measures for interface-based product evaluations. Empirical UX research relies heavily on self-report measures, such as questionnaires and interviews [1]. However, during such data collection methods, responses can be distorted, deliberately or unconsciously [3].

The IAT is a measure of association strength between concepts that relies on response latency [8]. It is believed to better reflect automatic attitudes than explicit measures and reduce the influence of self-report biases. It has also been shown to be resistant to deliberate faking [25].

The IAT is scientifically accepted as an implicit measure due to its high internal consistency, a rare occurrence for latency-based measures, and satisfactory test-retest reliability [23]. Additionally, the IAT is simple and fast to administer, requiring only a common PC.

We assessed the participants' aesthetic judgments of valence and self-identification towards interface images using the IAT and explicit measures. Interface aesthetics is believed to play a major role in users' perception of the system's usability [26] [29]. There's evidence in the literature indicating that aesthetic evaluations can occur in a spontaneous and automatic manner [21] [24]. We believe that the latency based nature of the IAT makes it an adequate method for measuring such dimensions.

In Section II, we present some of the limitations of self-report measures and how they led to the development of implicit measuring techniques. The IAT, the implicit measure used in this study, is described in Section III. In Section IV, we discuss how aesthetics can affect the perceived usability of a system. Section V talks about how judgments of aesthetic nature can be formed very quickly and why the IAT is an apt method to evaluate them. In Section VI, we describe the design and methodological implementation of the study, followed by a presentation of results, in Section VII. Finally, in Section VIII, we discuss how the results we obtained suggest that the IAT is a method that bears further study for UX research purposes.

II. SELF-REPORT LIMITATIONS AND IMPLICIT MEASURES

In a 2011 study [1], the authors found that two self-report methods - questionnaires and interviews - comprised more than half of the data collection methods used for empirical UX research. The authors, who reviewed 66 empirical studies from 51 publications, also found that the most assessed dimensions are emotions, enjoyment and aesthetics.

There's ample evidence that self-report techniques satisfy important psychometric criteria such as usefulness and efficiency. On the other hand, it is also known that they come accompanied by some limitations [3].

These issues include social-desirability and self-presentation motivations, which can make respondents purposefully distort their answers, particularly when these answers are believed to violate social norms, jeopardize one's self-image or go against the stereotypical answer [3].

Wrong answers can also be provided not deliberately but due to the respondents' inability to introspectively access previously formed attitudes [7].

In order to be able to infer mental contents while overcoming these limitations, researchers, mainly from the field of psychology, developed several techniques known as implicit measures [6].

These methods are called implicit because they aim to measure implicit attitudes, which, according to [7], are "*introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or* *unfavorable feeling, thought or action toward social objects*".

They're also implicit in the sense that they don't depend on the awareness of the participants relatively to what's being assessed with the procedure. Implicit measures are thus more sensitive to the spontaneous, automatic evaluations that can be assumed to guide real-life behavior and less likely to be influenced by factors like social desirability and self-presentation [4].

III. IMPLICIT ASSOCIATION TEST

The IAT, developed by Greenwald, McGhee and Schwartz [8] is probably the most well-known of all the implicit measuring methods.

The IAT indirectly measures the strengths of associations among concepts by requiring participants to sort stimulus exemplars from two pairs of concepts (e.g., Flower and Insect and Good and Bad) using just two response options [23].

The rationale behind the IAT is that sorting is facilitated (i.e., response latency should be lower) when two strongly associated concepts (e.g., Flower + Good) require pressing a response key and another pair (e.g., Insect + Bad) requires pressing the other response key. In contrast, when strongly associated concepts require a different response key (e.g., Insect + Good and Flower + Bad), sorting should be slower. The difference in response latency between these two tasks, called the IAT effect, is taken as an indicator of the degree of the strength of association between concepts.

As a measure of association strength between concepts, the IAT can potentially reveal different associations than the ones available introspectively and explicitly reported [14].

Among the factors that contribute to the IAT's acceptance as an implicit measure, are its high internal consistency, a rare occurrence for latency-based measures, and its satisfactory test-retest reliability [23].

It has also been shown that the IAT effect is resistant to several procedural artifacts. These include the hand assigned to each category, the variability in the number of items used to represent the concepts, the subject's familiarity with the items used to represent the concepts, the variability in the response-stimulus interval and the order of the mixed categorization task (as long there's counterbalancing of the order of the study) [3].

There is also evidence of the IAT's high resistance to faking when compared to self-report measures. For example, Steffens [25] found that the IAT, while not immune to faking, is much harder to fake than explicit measures.

Lucas and Baird [15] mention that potential self-report errors are unlikely to be shared across different measuring methods. They advocate a multi-method approach where self-report measures are complemented or validated by other techniques like implicit measures.

With this in mind, coupled with the fact that there aren't examples in the literature of the IAT being used in a UX research context, we aim to explore the potential of this method for interface-based product evaluations.

We assessed two non-instrumental dimensions regarding the evaluated products: valence and self-identification. In the valence IATs, interface pictures of the target products were paired with words pertaining to the attribute categories "Good" and "Bad".

For assessing the self-identification construct, i.e., the degree through which one associates itself with a given product, the image stimuli representing the target products were paired with words representing the attribute categories "Me" and "Others".

These are two constructs that the literature indicates the IAT is capable of assessing. According to Brunel et al. [3], "*the IAT can provide implicit measures of automatic attitudes, self-concepts, self-esteems, and stereotypes.*"

Since we wanted the evaluation to be based as much as possible on aesthetic judgments, no interaction with the products depicted by their interface pictures happened during the study.

In UX practice and research, it is common to design interfaces in non-interactive mediums (e.g., paper, digital images). The ability to evaluate non-interactive interfaces adds to the IAT's usefulness for UX purposes.

IV. AESTHETICS AND PERCEIVED USABILITY

There's evidence that interface aesthetics play a major role in perceived usability, influencing the users' perceptions regarding a system's ease of use not only before, but also after the user has interacted with the system.

This relationship between interface aesthetics and perceived usability has been demonstrated by Tractinsky [26]. The author conducted a study to validate and replicate a study conducted in Japan by Kurosu and Kashima that found that interface aesthetics play a major role in people's perceptions of apparent usability. The author was able to replicate the results in a different cultural setting (Israel), thus concluding that the relationship between perceived interface aesthetics and apparent usability is culturally independent. Moreover, the close relationship between these two dimensions also increases the likelihood that aesthetics may considerably impact system acceptability.

In a subsequent study, Tractinsky, Katz and Ikar [29] explored the relationship between interface aesthetics and usability both before and after users interacted with the system. In addition to corroborate that the users' perception of interface aesthetics is highly correlated with the system's perceived usability, the authors also found, to their surprise, that post-experiment perceptions of the system usability are influenced by the interface's aesthetics, not by the actual usability of the system.

As such, products more strongly associated with positive aesthetic valence might be perceived as easier to use and more easily accepted.

V. AUTOMATIC AESTHETIC JUDGMENTS

The IAT is a timed task which requires the participants to act as quickly as possible while avoiding making mistakes.

Thus, one can question if a limited exposure time to stimuli of variable complexity suffices for evaluative purposes.

 ISBN: 978-1-61208-325-4

We don't believe this to be an issue, due to the evidence indicating that consistent aesthetic evaluations can be formed rather quickly.

Lindgaard et al. [15] found that evaluations of the visual appeal of web homepages after only 50ms of exposure were highly correlated with judgments made after 500ms. The authors state that as little as 50ms might be enough for users to form a highly consistent aesthetic impression of a web homepage.

In a study aiming to replicate and expand Lindgaard et al. findings [15], Tractinsky and colleagues [28] found that aesthetic judgments could be formed after an exposition of only 500ms and that these perceptions are fairly stable, particularly in the case of extreme evaluations. The authors thus suggest that visual aesthetics plays an important role in user's evaluation of web pages and interactive systems in general.

Similarly, Locher et al. [16] found that ratings of paintings made after 100ms were highly correlated with ratings made after unlimited exposure.

Regarding the processes involved, Hekkert [11] believes that aesthetic pleasure or displeasure results uniquely from sensory perception and occurs at initial, mostly automatic and perceptual levels.

According to Norman's [22] framework, these automatic, sensory-based judgments are processed at the "visceral" level. Jordan [13] also speaks of pleasure originated by sensory perception, the physio-pleasure.

Tractinsky [27] speaks of the need to consider this "visceral" beauty as having a major influence on evaluations of beauty.

We believe that the IAT is an adequate tool to assess judgments of aesthetic valence, since it is believed to tap automatic processes.

This assumption is the behind some studies from the field of experimental aesthetics, which employed the IAT for assessing automatic aesthetic evaluations of paintings and architectural styles [21] [24], as well as visual patterns [20] [2].

VI. METHOD

A. Procedure

Eight participants (four female and four male) volunteered to participate in the study after being contacted by e-mail. All had normal or corrected to normal vision and were collaborators of the Fraunhofer Portugal AICOS institute, in Portugal. The average age was of 27.25 (SD = 4.06). The study took place at the institute's facilities. Participants were seated in front of a laptop running the Windows 7 operating system with a screen resolution of 1920x1080 pixels. A trial version of Millisecond's Inquisit software was installed in the laptop and used to administer the IATs and self-report measures.

At the beginning of each session, participants were asked to follow the on-screen instructions and given an opportunity to clarify any doubts. Each session took about 35 minutes.

The recommendations found in [23], regarding the IAT's structure and the need to counterbalance the order of the combined tasks and the presentation of measures were followed. Likewise, more than a single stimulus was used per category in order not to compromise IAT effect's magnitude and reliability.

B. Measures

1) IAT: All the IATs used in this study were based on the Picture IAT script available on Millisecond's Inquisit website.

Some adaptations were made, namely the translation of all the text displayed from English to European Portuguese, the participants' native tongue. The stimuli were also altered to fit the study's purposes.

Four pairs of target concepts were assessed: two mobile operative systems – Android and iOS; two Portuguese newspaper web sites – P3 and i Newspaper; two versions of an interface from a Web application in development – Prototype and Mockup; and two modifications of the Android mobile OS which provide a set of apps targeted at older adults – Smart Companion and Fujitsu. Figure 1 and Figure 2 show some exemplars of the image stimuli used in the IATs.

The order in which the IATs and corresponding explicit measures were presented to half of the participants is displayed in Table I. The other half completed them in reverse order.

Four IATs assessed the valence construct by pairing interface pictures of each target product with words representing the attribute categories "Good" and "Bad".

The self-identification construct was assessed in two IATs, where interface pictures from each product were paired with words from the categories "Me" and "Others".

TABLE I. ADMINISTERED IATS AND EXPLICIT MEASURES

Target concepts	Image stimuli	Attribute categories	Word stimuli
Android/iOS	6	Good/Bad	16
Android/iOS	6	Me/Others	10
Android/iOS valence and self-identification explicit measures			
P3/i newspaper	6	Good/Bad	16
P3/i newspaper valence explicit measure			
Prototype/Mockup	4	Good/Bad	10
Prototype/Mockup valence explicit measure			
Smart Companion/Fujitsu	8	Good/Bad	16
Smart Companion/Fujitsu	8	Me/Others	10
Smart Companion/Fujitsu valence and identification explicit measures			

The script used in this study computes the IAT scores using the improved algorithm developed by Greenwald, Nosek and Banaji [9]. No modifications were made to the scoring procedures.

Table I shows the number of stimuli used to represent the IATs target concepts (images) and attribute categories (words).

Even though the degree of familiarity of the participants with the concepts varied, this is a procedural factor the IAT is resistant to [3].

Figure 1. Exemplars of the image stimuli used in the Smart Companion/Fujitsu and Android/iOS IATs

2) Explicit measures: Two bipolar scales with five and seven points were used for assessing, respectively, the valence and self-identification constructs. Like with the IAT, the self-identification dimension was only assessed for two of the four product pairs.

The scale answers were coded from negative to positive values to represent a relative measure conceptually similar to the IAT D score (e.g., for the targets Smart Companion and Fujitsu, a score of -1 would indicate a moderate preference for Fujitsu relative to Smart Companion).

Since the order of presentation of measures was counterbalanced, half of the participants answered the self-report measures before the IAT. In order to provide reference to these users, one picture of each product's interface was displayed alongside the scales.

C. Stimuli

A total of 24 image stimuli representing the target concepts were used.

The images' width and height were normalized between IATs.

Since the degree of familiarity of the participants with the products varied, a cue was provided to reduce task confusion and facilitate categorization, as recommended by Brunel, Tietje and Greenwald [3].

The cue consisted of a label identifying the category membership of each picture stimuli placed below each image (e.g., images of the interface of Smart Companion had a label beneath saying "Smart Companion").

Labels were not used for the Android/iOS IATs, since these two concepts and the images used to represent them were assumed to be familiar to the users.

When translated from the original Portuguese used in the study to English, the items for the "Good" attribute category were "Joy", "Love", "Peace", "Wonderful", "Pleasure", "Glorious", "Laugh", and "Happy".

For the "Bad" attribute category, the following word stimuli were used: "Agony", "Terrible", "Horrible", "Bad", "Awful", "Failure", "Injured", and "Evil".

Regarding the self-identification IATs, the words for the attribute category "Me" were "Self", "Me", "Am", "Mine" and "I". For the attribute category "Others" they were "His", "Their", "Other", "Them" and "Others".

The words used for the attribute categories were mainly from Inquisit's default set, with some being adapted from literature examples.

D. IAT Design

Table II demonstrates the IAT structure used in this study.

TABLE II. IAT STRUCTURE

Block	Trials	Items assigned to left-key response [e]	Items assigned to right-key response [i]
B1	20	Android pictures	iOS pictures
B2	20	Good words	Bad words
B3	20	Android pictures + Good words	iOS pictures + Bad words
B4	40	Android pictures + Good words	iOS pictures + Bad words
B5	20	iOS pictures	Android pictures
B6	20	iOS pictures + Good words	Android pictures + Bad words
B7	40	iOS pictures + Good words	Android pictures + Bad words

A trial corresponds to the time that mediates from the display of a single stimulus until the correct categorization of that stimulus.

Every time an error was made, a red "X" was displayed below the stimuli and the participant was required to correct the error before proceeding to the next task.

The order of the combined tasks (pictures + words) was counterbalanced. Half of the participants completed the IATs in a similar order to the one above, while the other half completed it with the blocks B1, B3 and B4 switched with B5, B6 and B7.

The script used for the IATs in this study applies the improved scoring algorithm from Greenwald, Nosek and Banaji [9] when computing the IAT effect (D score).

Figure 2. Exemplars of the image stimuli used in the Prototype/Mockup and P3/i newspaper IATs

This improved scoring algorithm provides several benefits when compared to the previous version, namely: a higher resistance to the response speed artifact, almost eliminating the production of extreme scores for slow responders; a higher resistance to prior IAT experience (although not totally eliminating this effect); a better reflection of the underlying association strengths; a more powerful assessment of the relations between association strengths and other variables of interest; an increased power to observe the effect of experimental manipulations on association strengths; and a better insight regarding the individual differences that are due to association strengths rather than other variables.

There's one procedure that is not applied automatically by Millisecond's Inquisit software, which is the elimination of participants with response latency lower than 300ms in more than 10% of the trials.

Raw data was analyzed for every participant and no such cases were found.

VII. RESULTS

A. IAT

Table III shows the average D scores for the IATs. The interpretation of the IAT effect below each scored follows the conventional method: a D score lower or equal to 0.15 in absolute value is considered to indicate little no preference.

Absolute values between 0.16 and 0.35 and 0.36 to 0.65 correspond, respectively, to slight and moderate preference.

Values greater than absolute 0.65 are interpreted as indicating a strong preference.

The IAT effect was only noticeable in two cases: the Smart Companion/Fujitsu and Android/iOS valence IATs.

In both cases a slight preference was found for Smart Companion relative to Fujitsu (D = 0.33) and for Android relative to iOS (D = 0.30).

The remaining average D scores revealed no preference (D <= 0.15) for any of the target concepts.

TABLE III. IAT AVERAGE D SCORES

Target concepts	Valence	Self-identification
Smart Companion/Fujitsu	D = 0.33 (SD=0.38) Slight preference for Smart Companion	D = -0.09 (SD=0.43) Little to no preference
Android/iOS	D = 0.30 (SD=0.55) Slight preference for Android	D = 0.11 (SD=0.45) Little to no preference
P3/i newspaper	D = -0.07 (SD=0.51) Little to no preference	
Prototype/Mockup	D = 0.003 (SD=0.13) Little to no preference	

B. Explicit measures

Table IV displays the average scores of the self-report scales administered for each product pair evaluated. The scores were coded to directly map to the IAT's D score, providing a conceptually equivalent measure of relative preference.

Smart Companion was moderately preferred to Fujitsu on the valence scale (score = 1) and slightly preferred on the identification scale (score = 0.875). In all the other product pairs evaluated, no relative preference was reported.

TABLE IV. EXPLICIT MEASURES AVERAGE SCORES

Target concepts	Valence	Self-identification
Smart Companion/Fujitsu	Average score = 1 (SD=1.07) Moderate preference for Smart Companion	Average score = 0.875 (SD=1.73) Slight preference for Smart Companion
Android/iOS	Average score = -0.25 (SD=1.70) Little to no preference	Average score = -0.25 (SD=2.43) Little to no preference
P3/i newspaper	Average score = 0 (SD=1.20) Little to no preference	
Prototype/Mockup	Average score = 0 (SD=0.13) Little to no preference	

C. Data Correlation

1) IAT/Explicit measures: The average correlation coefficient between the IATs and explicit measures was 0.42 ($p > 0.05$), a relatively high correlation coefficient for implicit-explicit measures when compared to the average value of 0.24 found by Hofmann et al. [8].

Implicit and explicit measures tend to correlate when the evaluated objects are not socially controversial [8].

Since interface-based evaluation isn't a particularly sensitive topic, this outcome could be expected.

The correlations between the IAT and the corresponding explicit measure for each product pair evaluated are shown in Table V.

TABLE V. IAT/EXPLICIT MEASURES CORRELATION

Target concepts	Valence	Self-Identification
Smart Companion/Fujitsu	r = 0.12 (p > 0.05)	r = 0.40 (p > 0.05)
Android/iOS	r = 0.89 (p < 0.01)	r = 0.66 (p > 0.05)
P3/i newspaper	r = 0.56 (p > 0.05)	
Prototype/Mockup	r = -0.01 (p > 0.05)	

The lack of correlation in the Smart Companion/Fujitsu valence IAT leads us to hypothesize that it might have been a case self-report bias.

Smart Companion is an Android modification developed by Fraunhofer Portugal AICOS. All the participants in this study were collaborators of the institute and most of them were aware of this fact.

This might have led the participants to explicitly report a preference for the product developed by their employer instead of the competing solution.

After looking at each participant's scores we found that the majority (75%) explicitly reported a moderate to strong preference for Smart Companion. However, some of the individual implicit scores directly contradicted the provided explicit answers.

Summing all this up, it's a reasonable hypothesis that this might have been a case where the IAT scores reflect the method's resistance to self-report biases, thus resulting in a low correlation.

The existence of varying correlations between the IAT and explicit measures has been subject of ample discussion in the literature.

According to Brunel, Tietje and Greenwald [3], the correlation between IAT and explicit measures can be limited by several factors. These include factors related with self-presentation and social desirability concerns, poor introspective access to attitudes that may cause inaccuracy in self-reports and the existence of homogeneous attitudes across specific populations.

Hofmann et al. [12] refer that the IAT and explicit measures tap distinct constructs which can be more or less linked. According to Greenwald and Banaji [7], implicit measures can tap unconscious processes not accessible to explicit, and conscious, self-report. Thus, the correlation between the IAT and explicit self-report measures can depend of the correspondence between conscious and unconscious cognition.

They also mention procedural factors that can interfere with the correlations between measures, like the order of the IAT's combined tasks or the order of presentation of measures.

Even though methodological precautions were taken – including counterbalancing the order of the IAT's combined blocks and the presentation of measures – we can't rule out the possible interference of procedural factors.

2) Valence/self-identification IATs: On average, a strong correlation was found between the valence and self-identification IATs ($r = 0.70$; $p = 0.05$).

Hassenzhal's model of aesthetic experience [10] provides a potential explanation for these results. The author suggests that beauty is strongly related with the hedonic attribute of identification, which he describes as the communication of personally relevant values through objects. The correlation values for the Smart Companion/Fujitsu and Android/iOS pairs are displayed on Table VI.

TABLE VI. VALENCE/SELF-IDENTIFICATION IAT CORRELATION

Target concepts	Valence/Self-identification
Smart Companion/Fujitsu	r = 0.65 (p > 0.05)
Android/iOS	r = 0.75 (p < 0.01)

3) Valence and self-identification explicit measures: In order to see if the valence and self-identification explicit measures displayed a different relationship than the IATs, we also computed correlation coefficients for them.

On average, the valence and self-identification explicit measures were strongly correlated ($r = 0.82$; $p < 0.05$).

The individual correlation values can be found on Table VII.

TABLE VII. VALENCE/IDENTIFICATION EXPLICIT MEASURES CORRELATION

Target concepts	Valence/Self-identification
Smart Companion/Fujitsu	r = 0.70 (p = 0.05)
Android/iOS	r = 0.93 (p < 0.01)

The strong correlation between the valence and self-identification explicit measures parallels what happened with the IATs scores.

As mentioned previously, some authors believe that beauty is strongly related with the concept of identification, so a high correlation between these two constructs both for implicit and explicit measures doesn't comes as a surprise.

Moreover, this convergence between implicit and explicit measures can also be seen as indicative of the potential of the IAT as a complementary or substitutive method for self-report measures.

VIII. Conclusion

This study's main goal was to explore if the IAT could be a robust supplement or alternative to self-report measures for UX purposes, particularly for assessing non-instrumental aspects like valence and self-identification during interface-based evaluations. We believe our results suggest that possibility.

The main argument for our assumption is the fact that the IATs and self-report measures scores displayed a medium correlation ($r = 0.42$; $p > 0.05$). This value is substantially higher than the average correlation coefficient ($r = 0.24$) between the IAT and explicit measures found by a meta-analysis conducted by Hofmann et al. [8].

The assumption that the IAT can be used in parallel, or alternatively, to self-report measures is reinforced by the fact that a strong correlation was found both between the valence and self-identification IATs ($r = 0.70$; $p = 0.05$) and the equivalent explicit measures ($r = 0.82$; $p < 0.05$).

As such, these results indicate that the IAT was able to measure similar constructs to the ones evaluated through explicit reports in the context of this study.

Evidence from the literature indicates that the IAT taps spontaneous and automatic processes, which some authors believe that are responsible for guiding behavior [4] and judgments of aesthetic valence [11].

For interface-based evaluations with no interaction, where we assume that the aesthetic dimension plays a major role in the evaluative process, the IAT might represent more adequately user attitudes than explicit measures.

Another point in favor of the IAT is its resistance to unwilling or deliberate errors in self-report measures.

We hypothesize that there might have been a case of self-report bias in our study. Most of the participants knew that one of the products was developed by their employer, which might have led them to explicitly report preference for it relative to the other. However, the IAT didn't mirror those explicitly reported attitudes, resulting in a weak correlation between implicit and explicit measures.

Since there aren't, to our knowledge, previous studies using the IAT for UX research, we believe that this paper might contribute to pave the way for exploring this implicit measuring method as a complement or alternative to self-report measures, particularly for assessing non-instrumental qualities.

As future work, we aim to integrate the IAT in the testing and development cycles of Ambient Assisted Living (AAL) applications alongside self-report measures to assess non-instrumental qualities. By testing and following up with bigger user populations, our goal is to find if the IAT is indeed a more reliable way to collect data about user attitudes than explicit methods.

References

[1] J. A. Bargas-avila and K. Hornbæk, "Old Wine in New Bottles or Novel Challenges ? A Critical Analysis of Empirical Studies of User Experience," *Proc. 2011 Annu. Conf. Hum. factors Comput. Syst. - CHI '11*, pp. 2689–2698, 2011.

[2] M. Bertamini, A. Makin, and A. Pecchinenda, "Testing whether and when abstract symmetric patterns produce affective responses.," *PLoS One*, vol. 8, p. e68403, 2013.

[3] F. F. Brunel, B. C. Tietje, and A. G. Greenwald, "Is the Implicit Association Test a Valid and Valuable Measure of Implicit Consumer Social Cognition?," *Journal of Consumer Psychology*, vol. 14. pp. 385–404, 2004.

[4] J. De Houwer, "What are implicit measures and why are we using them?," in *Handbook of implicit cognition and addiction*, 2006, pp. 11–28.

[5] R. H. Fazio and M. A. Olson, "Implicit measures in social cognition research: their meaning and use.," *Annu. Rev. Psychol.*, vol. 54, pp. 297–327, 2003.

[6] B. Gawronski, "Ten frequently asked questions about implicit measures and their frequently supposed, but not entirely correct answers.," *Canadian Psychology/Psychologie canadienne*, vol. 50. pp. 141–150, 2009.

[7] A. G. Greenwald and M. R. Banaji, "Implicit social cognition: attitudes, self-esteem, and stereotypes.," *Psychol. Rev.*, vol. 102, pp. 4–27, 1995.

[8] A. G. Greenwald, D. E. McGhee, and J. L. Schwartz, "Measuring individual differences in implicit cognition: the implicit association test.," 1998.

[9] A. G. Greenwald, B. A. Nosek, and M. R. Banaji, "Understanding and using the implicit association test: I. An improved scoring algorithm," *J. Pers. Soc. Psychol.*, vol. 85, pp. 197–216, 2003.

[10] M. Hassenzahl, "The Interplay of Beauty, Goodness, and Usability in Interactive Products," *Human-Computer Interaction*, vol. 19. pp. 319–349, 2004.

[11] P. Hekkert, "Design aesthetics : principles of pleasure in design Design aesthetics : principles of pleasure in design," *Psychol. Sci.*, vol. 48, pp. 157 – 172, 2006.

[12] W. Hofmann, B. Gawronski, T. Gschwendner, H. Le, and M. Schmitt, "A meta-analysis on the correlation between the implicit association test and explicit self-report measures.," *Personal. Soc. Psychol. Bull.*, vol. 31, pp. 1369–1385, 2005.

[13] P. W. Jordan, "Pleasure with products: Human factors for body, mind and soul," in *Human factors in product design: Current practice and future trends*, 1999, pp. 206-217.

[14] K. A. Lane, M. R. Banaji, B. A. Nosek, and A. G. Greenwald, "Understanding and Using the Implicit Association Test: IV: What We Know (So Far) about the Method.," in *Implicit measures of attitude*, 2007, pp. 59–102.

[15] G. Lindgaard, G. Fernandes, C. Dudek, and J. Brown, "Attention web designers: You have 50 milliseconds to make a good first impression!," *Behaviour & Information Technology*, vol. 25. pp. 115–126, 2006.

[16] P. Locher, E. A. Krupinski, C. Mello-Thoms, and C. F. Nodine, "Visual interest in pictorial art during an aesthetic experience.," *Spat. Vis.*, vol. 21, pp. 55–77, 2007.

[17] R. E. Lucas and B. M. Baird, "Global self-assessment," in *Handbook of multimethod measurement in psychology*, 2005, pp. 29–42.

[18] S. Mahlke, "Aesthetic and Symbolic Qualities as Antecedents of Overall Judgements of Interactive Products," in People and

Computers XX - Engage. Proceedings of HCI 2006, 2007, pp. 57–64.

[19] D. Maison, A. G. Greenwald, and R. H. Bruin, "Predictive Validity of the Implicit Association Test in Studies of Brands, Consumer Attitudes, and Behavior," *Journal of Consumer Psychology*, vol. 14. pp. 405–415, 2004.

[20] A. D. J. Makin, A. Pecchinenda, and M. Bertamini, "Implicit affective evaluation of visual symmetry.," *Emotion*, vol. 12. pp. 1021–1030, 2012.

[21] S. Mastandrea, G. Bartoli, and G. Carrus, "The automatic aesthetic evaluation of different art and architectural styles.," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 5. pp. 126–134, 2011.

[22] D. A. Norman, *Emotional Design: Why We Love (Or Hate) Everyday Things*. New York: Basic Books, 2004.

[23] M. R. Banaji, "The Implicit Association Test at Age 7: A Methodological and Conceptual Review," *Test*, vol. 663. pp. 265–292, 2007.

[24] M. Pavlović, and S. Marković, "Automatic processes in aesthetic judgment: Insights from the implicit association test", *Psihologija*, vol. 45. pp. 377-393, 2012.

[25] M. C. Steffens, "Is the implicit association test immune to faking?," *Exp. Psychol.*, vol. 51, pp. 165–179, 2004.

[26] N. Tractinsky, "Aesthetics and apparent usability: empirically assessing cultural and methodological issues," in *CHI'97 Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, 1997, pp. 115–122.

[27] N. Tractinsky, "A Few Notes on the Study of Beauty in HCI," *Human-Computer Interaction*, vol. 19. pp. 351–357, 2004.

[28] N. Tractinsky, A. Cokhavi, M. Kirschenbaum, and T. Sharfi, "Evaluating the consistency of immediate aesthetic perceptions of web pages," *International Journal of Human-Computer Studies*, vol. 64. pp. 1071–1083, 2006.

[29] N. Tractinsky, A. . Katz, and D. Ikar, "What is beautiful is usable," *Interacting with Computers*, vol. 13. pp. 127–145, 2000.

Towards the Tangible Hyperlink

Alfredo Pérez Fernández
Department of Software
Engineering,
Norwegian University
of Science and Technology,
NO-7491 Trondheim, Norway
Email: perezfer@idi.ntnu.no

***Abstract*—Thanks to the persistent decrease in cost and size of electronics, this century is experiencing an important scale up in number of devices surrounding us, with a subsequent increment on the complexity of the user interaction with such devices. The proposed interaction concept - the tangible hyperlink - demonstrates that it is possible to apply the capacitive coupling communication technology to simplify the interaction with embedded devices without preventing users from maintaining control on their privacy.**

***Keywords*-CHI; Personal Area Network; Tangible User Interface; Privacy.**

I. INTRODUCTION

Abowd et al. [1], following Weiser's [2] vision, identified the requirement of "addressing some notion of scale [...] in *the number and type of devices, the physical space of distributed computing and the number of people using a system*". Thanks to the persistent decrease in cost and size of electronics, this century is experiencing an important scale up in number of devices surrounding us, with a subsequent increment on the complexity of the user interaction with such devices. Abowd et al. also proposed the request of providing "*natural interfaces* that facilitate a richer variety of communications capabilities between humans and computation".

Figure 1 illustrates the proposed concept: a *tangible hyperlink* that can be opened by physically "tapping" it. Information is transmitted bidirectionally to the wearable device through the skin of the user due to its capacitive property. This bidirectional transfer of information among devices through the recognition of the touch gesture entails both advantages and disadvantages. On one hand, it gives the freedom to develop such *natural interfaces* identified by [1], making it simpler to exchange information among embedded and wearable devices. On the other hand, a scenario where "devices and environments reveal not only personal information but also location and context information" [3] can make it incredibly complex for users to control their privacy. In this work, a novel concept of *tangible hyperlink* is presented as a mechanism to balance the usability of *personal area networks* (PAN) and the control that users maintain over their privacy. To achieve this balance, the key aspect is to use a metaphor (the hyperlink) to help users mimic the way they exchange tangible information to the way they exchange information on the web.

Figure 1: Concept of tangible hyperlink

The paper is structured as follows. In the next section, current related work is described. Section 3 describes the conditions in which the experiment was carried out. Section 4 provides the different results obtained from the experimental evaluation. The paper concludes with Section 5 with a summary and an outline of different possibilities for future research.

II. STATE OF THE ART

The first work presenting and demonstrating a device capable of communicating using an electrostatic capacitive coupled PAN is [4]. Based on this work, other authors such as Shinagawa et al. [5] propose different alternatives to improve the communication by means of using "electric-field sensor implemented with an electro-optic crystal and laser light". Hyoung et al. [6] study how different factors such as voltage, frequency and physical level codification affect communication performance. In contrast to most studies, Duck et al. [7] focus more on analyzing higher layers of the protocol stack communication. Hachisuka et al. [8] present a mechanism for simplification of modeling and fabrication of intrabody communication devices. They also analyze the impact on

 ISBN: 978-1-61208-325-4

cost/performance of such devices when using four electrodes instead of two. Ishii et al. [9] introduce their vision of *tangible bit* and emphasize the need of allowing users to *grasp and manipulate* the information in their surrounding as a form of tangible user interfaces, claiming that the goal is to “bridge the gaps between both cyberspace and the physical environment”. Based on this vision, Parkes et al. [10] introduce a prototype consisting of a “modular scalable building system with the physical immediacy of a soft and malleable material”. Glumes are able to communicate to each other and exchange information using electrostatic capacitive coupling and are designed to be responsive to human touch.

III. Experimental Settings

In this section, we describe an experiment carried out, in which a total of 19 participants were asked to use different devices designed to communicate to each other using electrostatic capacitive coupling and, afterwards, fill in a questionnaire. From the 19 participants, 12 were male and 7 female. 15 of the participants had some form of education on a technological field. The mean age was 25 years with a range from 20 to 42 years. The experiment was divided into two different phases and all the participants were involved in both of them. The first phase of the experiment consisted of an evaluation

Figure 2: The Tangible Greeting Application

of a very basic application, *the tangible greeting* (Figure 2). This first application had as a main target to capture the opinion of different users about the exchange of information through the touch gesture. The second phase of the experiment was the result of a redesign of the prototype based on the feedback obtained from the participants during the first phase. Due to the lack of control users had on the information they were providing to the system, a new concept was introduced: the *tangible hyperlink*. The *tangible hyperlink* served as a mechanism to acquire a compromise between usability and privacy by means of implementing a metaphor of something the participant was already familiar with, the web hyperlink.

1) First Phase: Tangible Greeting: The first phase of the evaluation consisted of a basic application displaying a personalized greeting message to the participants on a screen display.

Figure 3: The UTnP Bracelet

Each participant was asked individually to wear a *Universal Touch and Play* (UTnP) bracelet (Figure 3) to later touch a small metal surface connected to a PC. The *UTnP Bracelet*

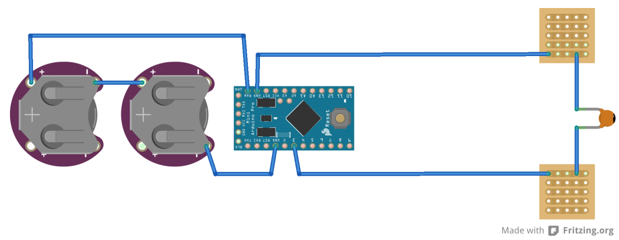


Figure 4: Implementation of the UTnP Bracelet

is implemented with an Arduino Pro Mini [11] (Figure 4) connected to a pair of electrodes. The electrodes are a pair of copper plates of 5.25 square centimeters, lying in parallel and separated from each other through a layer of a plastic insulator material (silicon and ethylene-vinyl acetate have been used without any noticeable performance decrease). The signal electrode is in contact to the skin of the participant, while the ground electrode is only in contact to the air. A ceramic capacitor of 1nF is used, connected in parallel, both to the signal and the ground electrodes, as proposed by [4]. The same configuration is applied on the signal receiver. Each

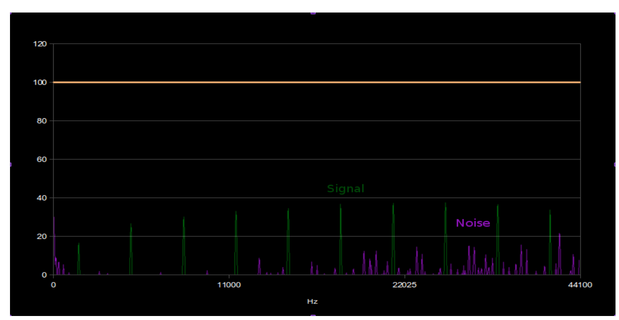


Figure 5: Frequency representation of received signal in percentage of magnitude

UTnP Bracelet generates a fixed *frequency-division multiplexing* (FDM) signal [12] that is decoded on the PC, after being transformed into the frequency domain. Figure 5 shows the received signal pattern, which ranges from 0.9Khz to 1.1Khz. The square wave generated from the ATmega produces a number of harmonic components after the base frequency. This redundancy is used as a mechanism to reduce the rate of false positives.

 ISBN: 978-1-61208-325-4

2) Second Phase: The Tangible Hyperlink: For the second phase of the evaluation, a change on the design was made. Instead of being the device of the participant the responsible for sending the information, it becomes a passive receiver, allowing the participant to decide on the amount of information that will be given away. For the implementation of the second

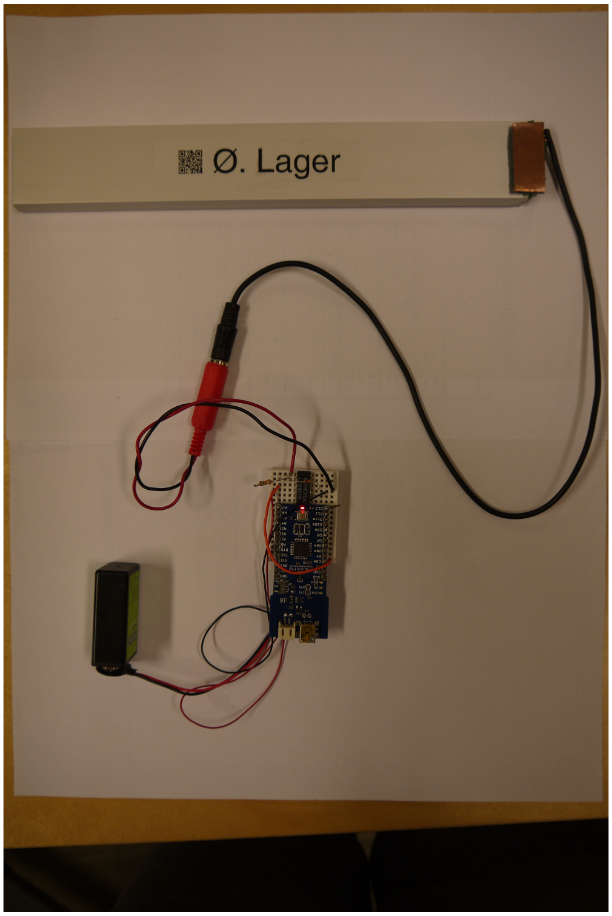


Figure 6: Name plate of a fictional person that encodes a tangible hyperlink

phase, an Android Smartphone located in an arm band was used in combination with the electrodes (Figure 1). The Android Smartphone was used to receive and process the signal generated from an Arduino Fio [13] embedded on a name plate (Figure 6). The participants were asked, invididually, to send a personalized greeting to a fictional person called Øyvind Lager. This task was performed by *clicking* on his *tangible link* (represented by a plate with his name). *Clicking* on the *tangible link* requests the participant to provide some personal information, including name, language and a greeting message. Participants were aware of the possibility of faking their identity or even cancelling the process at any time, which was a major difference comparing to the experiment of the first phase.

IV. Evaluation Results

After allowing the participants to familiarize and "play" within the scenario of each phase, they were given a questionnaire to fill in.

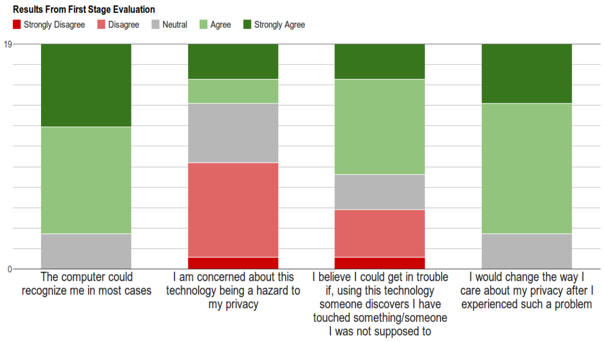


Figure 7: Results from questionnaires of first evaluation

First, it is possible to observe in Figure 7 that the prototype on the first phase behaved robustly, presenting little or no false positives or false negatives. 84.21% of participants stated that the computer was able to identify them. Something that can be interpreted as a paradox, from the second and third questions, is that 47.7% of participants disagreed with the statement that capacitive coupling can become a hazard to their privacy, while 57.89% agreed or strongly agreed that they could get in trouble if someone discovered that they had "touched something/someone they were not supposed to". This apparent paradox has its origin in the relative importance that participants give to this particular technology to their privacy. Participants stated that they were even more concerned with their privacy when using other technologies, including GPS, web browsing and social networks, because of their massive use.

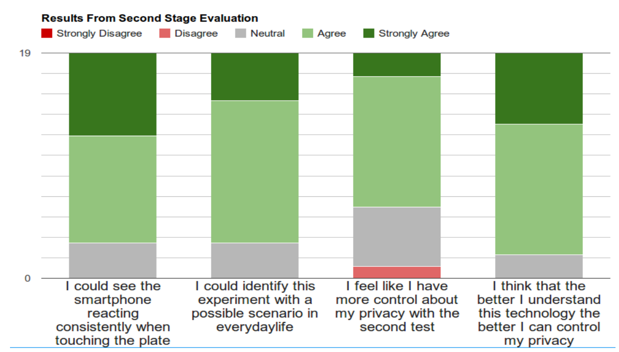


Figure 8: Results from questionnaires of second evaluation

The feedback obtained from the questionnaires of the first phase triggered a redesign of the prototype used in the second phase (Figure 8). Participants perception of the accuracy of the prototype was maintained, compared to the first phase. 68.42% of the participants claimed to have a better control on their privacy or the information they were providing with the second prototype. When asked which prototype they preferred, all the participants but three assured they preferred the last version, while the rest remained neutral. The results from the second phase suggest a qualitative improvement on the acceptance of the participants for those changes.

V. Conclusions and Future Work

Electrostatic capacitive coupling is an extremely powerful mechanism to share information among low powered embedded and wearable devices. It can enrich the touch gesture and make it become a much more meaningful component for many *pervasive applications*, but it can also become a hazard to privacy. The *tangible hyperlink* has been proposed as an approach to let the user have more control on the information that is provided to the system. It becomes highly important to incorporate metaphors to existing concepts, when designing new applications, so that it becomes easier for the user to have an understanding of the information flow. Arduino and Android devices have been used to implement a functional prototype for information exchange using electrostatic capacitive coupling. This setting is not optimal to achieve a high performance of data bandwidth and low power consumption but, instead, it offers a way to rapidly develop applications that need to adapt to very different and varying scenarios.

This study was carried out in the frame of a very basic application. The interest of the research will surely grow as the complexity of the scenario is increased. Further investigation is required to identify possible challenges and future findings. Part of this further investigation will consist of comparing this technology with other mechanisms of information exchange.

Acknowledgment

The author would like to thank for all the support received from the department of the software engineering group at NTNU.

References

[1] G. D. Abowd and E. D. Mynatt, “Charting past, present, and future research in ubiquitous computing,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 7, pp. 29–58, Mar. 2000. ACM ID: 344988.

[2] M. Weiser, “The computer for the 21st century,” *Scientific American*, vol. 265, no. 3, pp. 94–104, 1991.

[3] D. Anthony, T. Henderson, and D. Kotz, “Privacy in location-aware computing environments,” *IEEE Pervasive Computing*, vol. 6, no. 4, pp. 64–72, 2007.

[4] T. G. Zimmerman, “Personal area networks: near-field intrabody communication,” *IBM systems Journal*, vol. 35, no. 3.4, pp. 609–617, 1996.

[5] M. Shinagawa, M. Fukumoto, K. Ochiai, and H. Kyuragi, “A near-field-sensing transceiver for intrabody communication based on the electrooptic effect,” *Instrumentation and Measurement, IEEE Transactions on*, vol. 53, no. 6, pp. 1533–1538, 2004.

[6] C. H. Hyoung, J. B. Sung, J. H. Hwang, J. K. Kim, D. G. Park, and S. W. Kang, “A novel system for intrabody communication: touch-and-play,” in *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, pp. 1343–1346, 2006.

[7] D. G. Park, J. K. Kim, J. B. Sung, J. H. Hwang, C. H. Hyung, and S. W. Kang, “TAP: touch-and-play,” in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 677–680, 2006.

[8] K. Hachisuka et al., “Simplified circuit modeling and fabrication of intrabody communication devices,” *Sensors and actuators A: physical*, vol. 130, pp. 322–330, 2006.

[9] H. Ishii and B. Ullmer, “Tangible bits: towards seamless interfaces between people, bits and atoms,” in *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, pp. 234–241, 1997.

[10] A. Parkes, V. LeClerc, and H. Ishii, “Glume: exploring materiality in a soft augmented modular modeling system,” in *CHI’06 extended abstracts on Human factors in computing systems*, pp. 1211–1216, 2006.

[11] “Arduino Webpage- ArduinoBoardProMini.” [Online]. Available: http://arduino.cc/en/Main/ArduinoBoardProMini [retrieved February, 2014]

[12] J. F. Kurose and K. W. Ross, *Computer networking*. Pearson Education, 2012.

[13] “Arduino Webpage- ArduinoBoardFio.” [Online]. Available: http://arduino.cc/en/Main/ArduinoBoardFio [retrieved February, 2014]

Interaction With Mobile Devices by Elderly People: The Brazilian Scenario

Ricardo Roberto Leme, Luciana A. M. Zaina and Vitor Casadei
LERIS-Laboratory of Studies in Networks, Innovation and Software
Federal University of São Carlos
Sorocaba, Brazil
rleme@ufscar.br, lzaina@ufscar.br and vitor.casadei@gmail.com

***Abstract*—The use of mobile devices in Brazil has grown in different groups of people, making it an essential element in people's daily communication. Although there is a high concentration of users in the age group 20-40 years, it is the visible increase in the number of people over 60 that interacts with mobile devices, especially smartphones. These different groups of different ages have different needs with regard to interaction with mobile devices. The objective of this paper is to present the results of an exploratory study on the interaction of the older users with mobile devices in Brazil. For this, we used an exploratory study with potential users in Brazil. The exploratory study outcomes allowed us to draw the personas for the Brazilian scenario. The main interaction personas were identified and defined from this survey, aiming at an efficient representation of the group studied. The survey showed us that although the majority of users had mobile devices and used them for daily activities, some of them still face difficulties on handling them.**

Keywords-mobile devices; elderly user; personas.

I. Introduction

The proliferation of mobile devices has been increasing in recent years around the world. Currently, there are about 2.5 billion mobile phones in the world and about 80 % of the population has network coverage. It is estimated that in 2015 there will be over 5 billion active mobiles [1].

This phenomenon also occurs in Brazil. Among the most commonly used mobile devices can highlight the smartphone with the resources that are increasingly seeking to improve user interaction with the resources of the devices and their applications. However, the needs and preferences are diverse forms of interaction for each group of users, as elderly users.

According to the United Nations [2], there are currently nearly 900 million people aged 60 or more in the world. The expectation is that, by 2050, this number will reach 2.4 billion people, demonstrating that population growth among the elderly will be faster compared to other sectors of the population. In Brazil, this growth is no different. According to the 2010 Censo [3] prepared by IBGE (Brazilian Institute of Geography and Statistics), the population aged 60 or older, which was 4.8 % in 1991, came to 5.9 % in 2000 and reached 7.4 % in 2010. The institute is estimating that in 2025 the population of elderly people in the country will reach 34 million people. Evaluating the last decade, the number of elderly in the country increased by almost 4 million people, the result of vegetative growth and the gradual increase in average life expectancy. This is certainly a large group of people, with growth trend in the coming years.

The purpose of this article is to identify the characteristics of the users of old age and collect the main forms of use, difficulties and facilities in view of this group of users. The identification will be obtained through an exploratory study on the interaction of this group of users with mobile devices in Brazil. From this survey, we identified and defined the main personas interaction, aiming at an efficient representation of the group studied. As a contribution, it is hoped that this article demonstrates for developers of applications for mobile devices, the need to design specific forms of interaction for users of old age.

The paper is structured as follows: Section 2 presents the exploratory research and discussion of the survey results, Section 3 presents personas raised and Section 4 discusses some conclusions of the research.

II. Exploratory Survey

Users interact with computer systems, whether in desktop environment, web or mobile elements via its user interface (the part visible to the user of the system). This interface can be understood as a space of communication, one semiotic system in which signs are created and used to promote communication [4] [5]. The challenge of a user interface is to provide a space for communication so that this makes sense to the user who will use it [6]. During the searching for references on the subject, no articles were found about South America. Observing the lack in this area, the main goal was to collect data of the older user groups, analyzing their difficulties.

In this sense, the development of interfaces for mobile devices public-facing seniors should clarify the meaning that these users have in relation to the environment and the systems in which they interact in their day-to-day [7].

In order to achieve the goal, a questionnaire was formulated and given to the target audience. The planning and carrying out of the activity are described below.

A. Planning

Marconi and Lakatos [8] claim that an exploratory research can use questionnaires, interviews, participant observation and content analysis. In this work, we decided for the development and application on a questionnaire online; the application targeted for users participating in two groups of elderly. The content of the questionnaire was assembled based on related work and the goal set. The questionnaire has two sections, namely, the first on education level, age group and sex, and the second one on the use of mobile devices for seniors. In the survey on the age of the subjects, we planned to include users from age 50 to survey trends for the next decade. Although there are changes in the knowledge given in relation to technological developments, it is clear that users of old age usually fail to follow this evolution. If in the past the difficulty was the interface of a VCR (Videocassette Recorder) tape, now is the interface of a smartphone, which can be anticipated here for 1 or 2 decades? With this, it is observed that age exerts a strong influence on the use of information technology and communication in which there is a significant reduction from the 50 [9]. Nine (9) questions were prepared, and one (1) open ended question was and the other multiple choice.

Established the goal, hypothesis, methodology and materials used for the verification of the object of study:

- Hypothesis: Considering the users' interaction with the elderly mobile devices, there are many difficulties or points that could be improved to meet the needs of this user group.
- Objective: Collect the main forms of use, difficulties and facilities in the view of users of elderly facing the operationalization of mobile devices.
- Methodology: The proposed map the use of users of old age and thus define the personas, the user group was invited to participate in the completion of a questionnaire survey of socio-cultural and the use of mobile devices.
- Handout: To conduct the exploratory experiment we prepared a Questionnaire Survey Profile containing issues of social and cultural so that you can trace profiles of those users.
- Target Audience: Elderly users of social network Facebook to participate in groups.

B. *Execution of the Survey*

The presence of elderly users in the software market is still modest; considering the participation in Brazil of elderly users (aged 60 or older) on Facebook, for example, it is lower than 1% [10]. However, in the past few years there has been a growth in participation, and considering the current users aged 50 or above, which will be part of this group over the next decade, there will be a participation of more than 4%. Although this percentage does not seem expressive, it will represent over 2.8 million users [10] just in Brazil. Through the social networking site Facebook, an explanatory message was posted on the research and inviting the administrators of 41 groups of elderly in different cities of the country to publish the link of the research groups of closed social network.

The phase of contact with the administrators of groups of seniors, it was a very positive surprise, where the vast majority talked about the importance of research with this group of users. Only administrators of two groups (one from a local authority and other social network) showed resistance in the questionnaire. To these, we performed a more detailed explanation of the real goals.

The questionnaire was also applied to two elderly groups located in the city of Itu in Brazil with 8 people filling it manually. The link to the questionnaire has been accessed 297 times and got a total of 263 responses. For mapping the number of clicks, we used the Google URL Shortener [23] and the electronic form was created through Google Drive [24].

C. *Results and Discussion*

In the consolidation of the results, we considered both questionnaires in groups of elderly as the electronic questionnaire, totalizing 271 participants.

Understanding the educational, social and cultural issues assists in the process of formalizing the characteristics of the user group to be studied [8]. With this, we have analyzed the responses of users from the perspective of their profiles. The results are shown below.

The majority (98%) of respondents responded to the survey via the electronic form available on the Internet; 63% were female; 91% have cell phone.

For many users, one smartphone can become the only way to connect to the Internet [11], and, with it, the same way one can use a desktop PC, several software interface elements, completely different between each application, can be installed. Thus, it was raised from the research subjects, if besides their cell, how many % were smartphones. The survey shows that 149 subjects (56 %) have a smartphone, while in male subjects the number of people who have a smartphone reaches 72 %. The age range of respondents is 14 % between 51-59 years, 74 % between 60-75 years, 12 % between 76-85 years, and 0.01% or only 3 users older than 85 years.

Figure 1 shows that a majority of respondents (54 %) had higher education is "full" or "Post Grad". However, according to the questionnaire, 8 % of people have only the "elementary school incomplete", demonstrating that mobile devices are used by people with different levels of education; since, the motivation in creating interfaces considering the experience of non-specialist users of different ages and levels of knowledge [6].

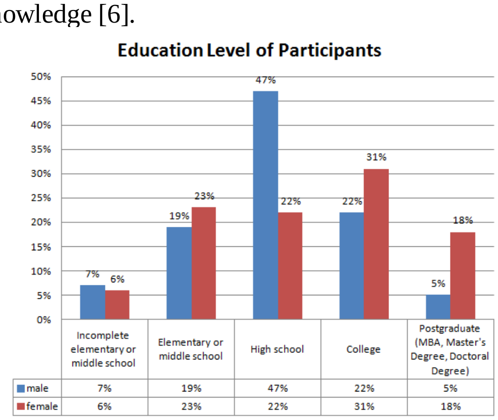


Figure 1 - Level of study by age

The users were asked about the main difficulties in the use of mobile devices (more than one option could have been selected). The highlight was the difficulty of visualizing the interface elements, where 58% of users above 76 years showed that ”the letter of the text is too small”, while 37% mentioned that ”I have trouble seeing the images of the cell”. The complete survey of the difficulties can be seen in Table 1.

TABLE I. MAIN DIFFICULTIES ENCOUNTERED IN THE USE OF MOBILE DEVICES BY ELDERLY PEOPLE

Difficulty Pointed	***To 60 years***	***From 60 to 75 years***	***Above 75 years***
The letter of the text is too small	12%	38%	58%
I have trouble seeing the images of the cell	8%	21%	37%
I find it hard to click on the buttons	4%	18%	29%
The programs are very complicated to use	17%	23%	41%
I do not have any difficulty	48%	22%	11%

Subjects above 60 years showed ”having no difficulty” in the use of mobile devices; 72 % are ”Higher” or are ”Post Graduates”, demonstrating a trend where the level of education can directly influence the difficulty of the operation of the interface.

The survey in Table 1 corroborates the set of challenges proposed by Nilsson [12] in developing user interfaces for mobile devices, where a major challenge is the efficient use of screen space, through the using different layouts that are appropriate to the device, mechanisms for grouping the information, screen scrolling, management of dialogues with the user from keyboards to touch screens resizing and management interface for different devices and types of users.

Considering a qualitative research, a essay question was included on other difficulties encountered in mobile devices, where tissues were some comments: ’*The configuration program has hidden options*’; ’*I find it interesting that the cell facilitates the lives of elderly since the technology is a permanent thing in our day-to-day.*’. ’*Why interaction with new technology is so complex?*’.

In order to verify the actual use that elderly users make of with their mobile devices, they were asked to talk about besides, what other interactions were performed on the devices (more than one option could have been selected). The result of the survey can be seen in Figure 2.

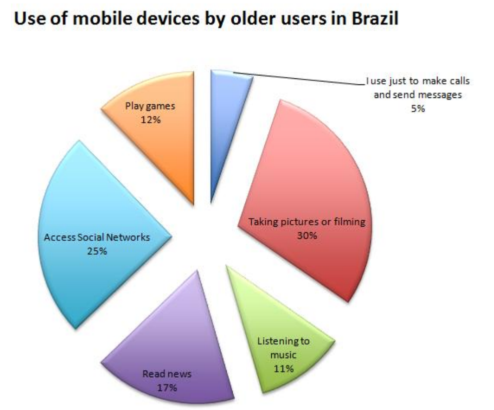


Figure 2- Use of mobile devices by elderly users

Aiming to raise new possibilities of development for mobile devices (smartphones and tablets with touch interaction), there were three questions, as shown in Table 2, in which the subject can respond if ”not would be interested”, ”would some interest” or ”Had a lot of interest” in each item.

TABLE II. POSSIBILITIES FOR NEW INTERACTION IN MOBILE

Activity	***No Interest***	***Interest***	***Very Interest***
Read updates from mobile phone only moved his hand	42%	53%	5%
Instead of writing the message, can dictate the phone the desired text	24%	30%	46%
Enable the cell to read aloud updates mobile	53%	37%	10%

It was observed that, according to Table 2, 77 % of the subjects had ’Interest’ or ’Very Interest’ to use the drive in the device reading the update messages mobile device, and 76 % shown ’Interest’ or ’Very Interest’ for “can dictate text instead of writing the message”. Only the activity Enabling “mobile reading aloud updates cell” had a majority group ’No Interest’ totaling 53 %. This item demonstrates that the development of a mobile application in order to reach all audiences, should also consider the needs and desires of a specific audience, such as the elderly people. In Brazil, about 80% of Internet aged users have education level above the national average, and economic classes are more favorable [13]. Thus, it is critical that developers of mobile applications consider the specific needs of this group of users, since the declines coming from natural aging can affect the users’ interaction with the interfaces.

Based on the motivation of this paper, a comparison of the results of this study was carried out conducted in Brazil, South America with information in United States, North America [21][22]. The result can be seen in Figure 3.

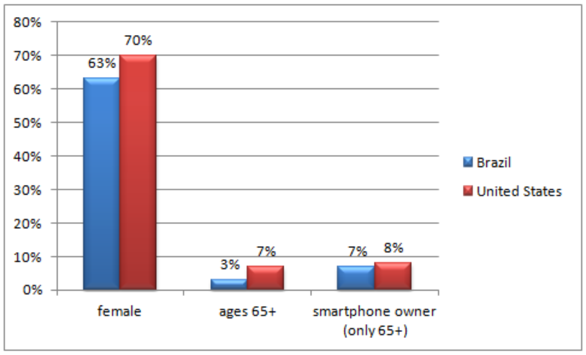


Figure 3 - Comparison of the results

III. IDENTIFICATION AND DEFINITION OF PERSONAS

The persona is defined as a fictional character, a hypothetical model of a group of real users, created to describe a typical user [14][15]. Personas have been used as an important mechanism for the precise definition of the characteristics of potential users of a system [16]. To identify and define the personas, one must first observe the profile of the user group being studied.

In this work, we decided to use the approach proposed by Nielsen [14] for modeling and analysis of user profiles. The personas are defined primarily by their goals, determined in a successive refinement process during the investigation of the domain and the characters that are part of it.

Cooper [15] directs that instead of organizing features to accommodate all people, should design the interface for a single persona. Although personas are fictional, there are strictly defined details to represent typical users. The method of identification and preparation of personas, adapted from [14], consisted of the following steps:

- Discovery of users: Through exploratory study, the questions were answered as 'Who are the users?' and 'What do they expect from an interface?'
- Construction Hypothesis: Through the tabulation and analysis of the exploratory study, we identified groups of related users.
- Searching for patterns: Through the categorization of the data, it was verified that all groups have proposed the hypothesis of equal importance.
- Construction of personas: From the Categorization of the study group personas were prepared.

The extraction of information from the database of users was performed using datamining software open-source Weka [20]. This software is used to identification of data clusters using techniques classification, clustering and association rules. Through Weka, using the KNN classifier (K - Nearest Neighbors) it was possible to identify the user profile collected on the various characteristics, since this demand K classifier elements of the data set that are closest to this unknown element, or which have the smallest distance. For this database was used Euclidean distance.

After collecting all the data, a new challenge arose which was to transform the entire mass of data into useful information and sufficiently well structured for the creation of personas. Thus, it is justified to use the Weka [20] which were identified by cross-checking logic of user segmentation that will define personas profiles. From the questionnaire, the preferences of interaction, economic profile, education, personal goals and practical goals were observed.

With this analysis, we reached at three personas featuring the group studied. Figures 3, 4 and 5 illustrate the personas encountered.

Figure 4 - Persona 1 (Manoel) obtained in the survey

Figure 5 - Persona 2 (Marina) obtained in the survey

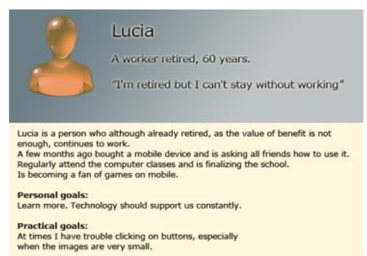


Figure 6 - Persona 3 (Lucia) obtained in the survey

Personas found from the exploratory study are quite similar to three groups identified by Gatto [17]. In his work, the author has identified five different types of personality of the elderly:

1) Constructive - well integrated, respected, stable, enjoying what life gives you;

2) Dependent - is passive, disengaged voluntarily and satisfied, sir rocking chair, because, finally, can rest;

3) Defensive - assets, rigid, disciplined, individualistic, dedicated to many activities because he could not stand still.

4) Choleric - blame the world and people for their personal failures, has little ambition for the future, social and economic standards unstable precarious; fight against the manifestations of aging;

5) Pessimistic - shows steady decrease in socioeconomic status and no history of life, hates himself, is depressed and isolated and often exaggerate their lack of physical and psychological training, becoming a victim. Accepted the sad

old age, but has death as its release this unsatisfactory existence.

The first three types (Constructive, Defensive and Dependent), each in their different lifestyle and worldview, have adapted successfully to the aging process. On the other hand, the last two types (Choleric and Pessimist) did not fit the aging [17].

Table 3 shows who represent these personas.

TABLE III. MAPPING OF PERSONAS RAISED

Persona	***Scholarity***	***New Interactions***
Manoel	Graduated	Texts of the programs may be higher
Marina	Basic Education	Afraid to use a cell
Lucia	High School	Images may be higher

From Table 3, three very different profiles personas should be considered in the development of interfaces for mobile devices. The persona has Manoel higher level, there is great difficulty in the operationalization of mobile devices, but due to his age, he has some discomfort in reading texts; since, he considers that the font size could be larger. The Lucia persona is the typical user of seniors who are having their first contact with the technology and uses the mobile device as well as a platform for fun. Already persona Marina is a user who should have special attention from the development team of new interfaces, because, as she has strong reluctance with the technology of mobile devices, any negative experience with the interface can get her to give up time to have this contact.

Another step in the construction of personas according to Nielsen [14] is validating personas created. The characteristic of this work, this activity was not performed, nor were inserted pictures of real people in the personas created as proposed by Nielsen [14] in order to protect copyright of the photographer and the photographed person.

IV. CONCLUSION

This article aimed at studying users of mobile devices for the elderly in Brazil. Thus, the results represent trends, experiences and behavior patterns associated only with the study group. The cultural, religious, political, technical, can not be determined by the number of subjects studied. However, research can be considered as an initial study to help understand the use of mobile devices by users seniors and opportunities for future research, development or adaptation in the software specific for this group, since the software must serve the needs users [18] [19].

From this study, it is clear that an application of global reach such as social networks or solutions of e-commerce cannot consider, for example, only users of a certain age, because the interests and needs of each band is totally different, as was found in accordance with the personas raised.

The research showed that, beyond the limitations of motor, cognitive or perceptual common users of old age, there are other obstacles in the use of mobile devices in the user group of seniors. The difficulty in performing operations theoretically simple as reading a text or click a picture, suggest the need for targeted approaches to the public studied.

Thus, it is concluded that the need to understand how the techniques and methodologies aimed at IHC group of seniors can help to understand the real needs of this group of users, the social aspects and experience with technology users can interfere with successful acceptance of new software or technology.

ACKNOWLEDGMENT

The authors wish to thank Centro Paula Souza and the State Government of Sao Paulo for the financial support to this study.

REFERENCES

[1] G. Frattini, F. Ceccarini, F. Corvino, I. De Furio, F. Gaudino, P. Petriccione, R. Russo, V. Di Carlo, and G. Supino, “A new approach toward a modular multimodal interface for pdas and smartphones,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 5188 LNCS, 2008, pp. 179–191.

[2] United Nations, “Major rise in world’s elderly population: DESA report,” [retrieved: December, 2013]. [Online], Available: http://www.un.org/en/development/desa/news/population/majorrise-in.html.

[3] IBGE, “Censo 2010,” Website, 2011, [retrieved: December, 2013]. [Online], Available: http://censo2010.ibge.gov.br/.

[4] O. L. de Oliveira, “Design da interação em ambientes virtuais: uma abordagem semiótica,” Biblioteca Digital da Unicamp, 2000.

[5] C. S. de Souza and C. F. Leitão, “Contribuições da engenharia semiótica para os métodos científicos de pesquisa em ihc,” in Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems, ser. IHC ’08, 2008, pp. 348–349.

[6] A. M. Melo, M. C. C. Baranauskas, and S. C. d. M. Soares, “Design com crianças: da prática a um modelo de processo,” Revista Brasileira de Informática na Educação, vol. 16, no. 01, 2008.

[7] J.-H. Lee, E. Poliakoff, and C. Spence, “The effect of multimodal feedback presented via a touch screen on the performance of older adults,” Haptic and Audio Interaction Design, Proceedings, vol. 5763, 2009.

[8] M. d. A. Marconi and E. M. Lakatos, Research techniques, planning and execution of research, sampling and research techniques, preparation, analysis and interpretation of data, 7th ed., 2008.

[9] R. Hellman, “Universal design and mobile devices,” in Universal Acess in Human Computer Interaction. Coping with Diversity, ser. Lecture Notes in Computer Science, C. Stephanidis, Ed. Springer Berlin Heidelberg, 2007, vol. 4554, pp. 147–156. [retrieved: December, 2013]. [Online], Available: http://dx.doi.org/10.1007/978-3-540-73279-217

[10] Social Bakers, “User age distribution on Facebook in Brazil,” [retrieved: October, 2013]. [Online], Available: http://www.socialbakers.com/facebook-statistics/brazil.

[11] D. C. Robbins, B. Lee, and R. Fernandez, “Tapglance: designing a unified smartphone interface,” in Proceedings of the 7th ACM conference on Designing interactive systems. New York, NY, USA: ACM, 2008, pp. 386–394. [retrieved: December, 2013]. [Online], Available: http://doi.acm.org/10.1145/1394445.1394487

[12] E. G. Nilsson, “Design patterns for user interface for mobile applications,” Advances in Engineering Software, vol. 40, no. 12, 2009, pp. 1318 – 1328. [retrieved: December, 2013]. [Online],

Available:http://www.sciencedirect.com/science/article/pii/S0965997809000428

[13] CETIC.br, “Survey on the use of information and communication technologies - ict households and enterprises 2011,” [retrieved: December, 2013]. [Online], Available: http://op.ceptro.br/cgi-bin/cetic/tic-domicilios-eempresas-2011.pdf.

[14] L. Nielsen, Personas. Aarhus, Denmark: The Interaction Design Foundation, 2013. [retrieved: December, 2013]. [Online], Available: http://www.interactiondesign.org/encyclopedia/personas.html

[15] A. Cooper, R. Reimann, and D. Cronin, About face 3: the essentials of interaction design. Wiley, 2012.

[16] Y.-N. Chang, Y.-K. Lim, and E. Stolterman, “Personas: from theory to practices,” in Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges, ser. NordiCHI ’08. New York, NY, USA: ACM, 2008, pp. 439–442, http://doi.acm.org/10.1145/1463160.1463214.

[17] I. d. B. Gatto and M. P. Netto, “Aspectos psicológicos do envelhecimento; psychological aspects of aging,” Atheneu, 1996.

[18] X. Bao, M. Gowda, R. Mahajan, and R. R. Choudhury, “The case for psychological computing,” Hot Mobile - Pervasive Computing, vol. 12, no. 3, 2013, [retrieved: December, 2013]. [Online], Available:http://www.hotmobile.org/2013/papers/full/6.pdf.

[19] J. Feng, M. Johnston, and S. Bangalore, “Speech and multimodal interaction in mobile search,” in Proceedings of the 20th international conference companion on World wide web, ser. WWW ’11, 2011, pp. 293–294.

[20] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1, 2009, pp. 10-18.

[21] Nielsen, “State of the Media: The Social Media Report 2012” Website, 2012, [retrieved: December, 2013]. [Online], Available: http://www.nielsen.com/us/en/reports/2012/state-of-the-media-the-social-media-report-2012.html.

[22] comScore, “2013 Mobile Future in Focus”, Website, 2013, [retrieved: December, 2013]. [Online], Available: http://www.comscore.com/Insights/Presentations_and_Whitepapers/2013/2013_Mobile_Future_in_Focus.

[23] Google Inc., “Google URL Shortener”, Website, 2013, [retrieved: December, 2013]. [Online], Available: http://goo.gl/

[24] Google Inc., “Google Drive”, Website, 2013, [retrieved: December, 2013]. [Online], Available: http://drive.google.com/

A Set Of Heuristics for User Experience Evaluation in E-commerce Websites

Laia Bonastre, Toni Granollers
University of Lleida
Lleida, Spain
laia.bonastre@gmail.com, antoni.granollers@udl.cat

***Abstract*—Electronic Commerce (e-commerce) websites need to provide customers with a positive User Experience (UX) to be successful and competitive. In this paper, it is presented a set of 64 heuristics as a tool to evaluate the grade of UX achievement of these kinds of sites. The set is based on three studies which provide functional requirements and guidelines in regard to the quality of e-commerce web sites. The main contribution of this work is the standardization of these recommendations by formulating them in interrogative sentences to facilitate the evaluation of e-commerce sites. Each heuristic is accompanied by examples and suggestions that facilitate their evaluation by a provided scoring system.**

Keywords-Heuristic evaluation; E-commerce; User Experience.

I. INTRODUCTION

Business-To-Consumer (B2C) websites are complex interactive systems whose primary goal is to draw a visitor into completing an online purchase. It is well know that User eXperience (UX) is one of the most critical factors for Electronic Commerce (e-commerce) success and as such, it has become a competitive requirement. The standard ISO DIS 9241-210:2008 [7] gives one of the most popular definitions, it says that UX is: “A person's perceptions and responses that result from the use and/or anticipated use of a product, system or service”. Morville [14] has contributed in the field of UX providing seven well-known facets, these are: useful, usable, findable, valuable, desirable, accessible and credible.

However, it seems that UX is not as implemented in e-commerce sites as it should be. According to Baymard Institute [1], the average shopping cart abandonment rate from 18 different studies is 67%. Therefore, enhancing the UX is required in B2C sites as this in turn contributes to an increase in the number of orders. The first step to improve the UX in existing and new e-commerce websites is their evaluation.

One of the most used evaluation inspection methods is the Heuristic Evaluation (HE). It can be applied during the development process of interactive systems or used on real operational systems. The HE is an effective method to review interfaces by taking the recommendations based on User Centered Design rules and contrasting them with the website. These recommendations come in different guises, such as design principles, heuristics, guidelines, best practices or user interface design patterns and standards [10] that are supposed to serve interface designers and evaluators. Despite this, only heuristics are usually orientated as an evaluation more than a design tool. This is because they are often written in interrogative sentences instead of being written in declarative sentences. An study presented by Masip et al. [11] reveal that evaluators prefer interrogative sentences because they are more intuitive.

Another difficulty of HE is that it is necessary to adapt the heuristic set to the specific features of each interactive system. Thus, evaluators have to combine different recommendations sources in order to review the specific application domain. This involves a long reviewing process as guideline collections inevitably induce conflict between various resources. Guidelines conflicts are caused by the different characteristics of each set of recommendations. For instance, level of detail, format, scope, language, classification of the principles, updating, level of validity and quality and development phase for use [16]. In addition, each set uses different scoring systems to score the recommendations, so it is necessary to adjust it in the resultant set.

The process of combining different heuristic sets usually finishes up with a long list of duplicated recommendations, similar statements using different terms and potentially conflicting guidelines. A clean-up and selection process is then required to provide a reliable, consistent and easy to use heuristic set [9]. Furthermore, the majority of the recommendation sets are written in declarative sentences. The evaluators can keep to this way of writing, as it eases the adaptation process of the heuristic set, but then the implementation of the evaluation process will be a far more complicated task.

For e-commerce websites the adaptation of the heuristic set is especially challenging. It is possible to use universal usability principles, such as Nielsen's [15], but they do not cover all the aspects involved in B2C sites. Additional heuristics appear to be needed to support e-commerce specific components and functionalities, such as shopping cart, customer service, checkout and registration process, category and product pages. Moreover, there are some factors that are especially important for the customer, such as trustworthiness, safety and privacy.

Some specific e-commerce user interface recommendations are available; for example, the Bauer's guidelines for Product Pages [2] or the Guidelines of E-

 ISBN: 978-1-61208-325-4

Commerce Checkout Design presented by Holst [6]. As of the two previous examples, most of the sets of recommendations found for e-commerce websites are partial, presenting principles only for some aspects of B2C websites, such as shopping cart, product pages, checkout process, etc. In addition, any recommendations written in interrogative sentences have been found in this field.

For all the reasons above, it is proposed to pull all the e-commerce principles from the literature together and standardize them into interrogative sentences to be ready for evaluators to use. A scoring system is also suggested to ease the reviews and overall scoring of B2C sites. The resultant heuristic set will prevent the problems that arise from the combination of the recommendations taken from different sources. The aim of this paper is to begin with the recommendations standardization by providing ”a first set of heuristics“ to evaluate the UX in e-commerce websites.

This paper is structured as follows: first, in the methodology section, the procedure to obtain the heuristic set and the scoring system are explained. Then, in section 3, the resultant set and its detailed organization and presentation are shown. Finally, in the conclusion section, the main conclusions and the future work are shortly described.

II. Methodology

The heuristic set introduced is based on the analysis of three studies, these studies present guidelines or functional requirements for e-commerce websites. The selected studies are Liang and Lai [8], Lightner [9] and Cao et al. [3]. They were selected due to their similarity as their criterion is used to test on-line bookstores. These studies do not cover all the aspects of User Experience but as stated, this paper's aim is just to begin the standardization.

The first thing to do to obtain the proposed heuristic set was to deduce and recollect all the recommendations given by the three studies. Then, each recommendation was rewritten formulating an interrogative sentence and providing examples or advice to facilitate their scoring. In table I, an applied example is shown in order to clarify the process followed to obtain the heuristics. In the first column there are cited the three studies. In the second column there are quoted the author's definitions of their own recommendations. In the next column there are the guidelines about security extracted from the studies. Finally, in the last column, the resultant heuristics which are inspired by the three authors' recommendations are presented.

TABLE I. PROCESS FOLLOWED TO OBTAIN THE HEURISTICS

Author	**Authors' definition of their own recommendations**	**Recommendations about security from the studies**	**Resultant Heuristics**
T. P. Liang, and H. J. Lai [8]	“Functional requirements derived from the customer's perspective”	“I trust the web site will not misuse my personal information”	*Does the website show security logos?* It should show them especially in the checkout process to build trust. The logos may be related to the shipment (FedEx, UPS, etc.), payment options (Visa, PayPal, etc..) or security, like SSL.
		“The web site is secure”	
		“The web site is realiable”	*Does the website inform of the level of security when paying by credit card?* It should demonstrate that it is a safe website to buy from.
		“I fell protected/safe when I use the site”	
N. J. Lightner [9]	“Functional requirements that represent facets of customer service in a B2C site”	“Provide secure payment”	*If personal information is required by the website, does it have the Privacy Policy available?* It should include all possible uses of their personal information. The footer is the standard location to link this information. It has to be easily understandable by the customers.
		“Inform customer of paymen security (Safe Shopping Guarantee)”	
		“Inform customer of privacy policy (For instance: Link during checkout or help, Privacy icon link, etc.)”	*Does the website has safety certificates granted by external companies?* It is important to get them and let customers know about them. These are companies like VeriSign or ControlScan. It must use the secure SSL technology as well.
M. Cao, Q. Zhang, and J. Seydel [3]	“A set of factors that capture the quality of an e-commerce website”	*“Claiming security of transactions:* Since the seller and the buyer do not actually see each other in the virtual market, Internet security is a major concern. Therefore, adopting proper security mechanisms, such as secured electronic transactions, and making announcements may help.”	

Following that, the heuristics were organized based on the stages of customer's buying behavior described by Engel et al. [5]. Consequently, as it is explained in the next section, the heuristics are organized in the following six parts: Need Recognition and Problem Awareness, Information Search, Purchase decision making, Transaction, Post-sales Services and it has been added a section to gather factors that affect UX during the whole purchase process. This organization intends to facilitate the evaluators' understanding of customer behavior to enhance the HE results. If during the reviewing of the interface the evaluators follow the stages of the purchase decision process they will understand the customer's needs and point of view.

Finally, a scoring system has been established as a convenient tool to assist the evaluators in conclusion making by providing quantitative results of the HE. Each heuristic was assigned either a “yes” or “no” answers or a 1 to 5 Likert scale. To obtain a global score for the evaluated websites the affirmative answers get 5 points and the negative answers get 1. For the other system, the evaluator can grade the level of accomplishment between 1 to 5. The scale is represented as: 1 strongly disagree, 2 disagree, 3 neither agree nor disagree, 4 agree and 5 strongly agree. The first scoring system is assigned to the questions that either have or do not have specific features or functionalities. The Likert Scale system is attached to the heuristics that require a ranged score as they depend on expert opinion. Eventually, it is possible to combine scores together to obtain a total score out of 233 points. This eases the comparison between different websites, and is useful to estimate the level of UX that a website has according to the heuristic set.

III. Results

In this section, a set of 64 heuristics ready to be rated by experts is presented. The set is the result of the compilation and standardization of the recommendations from the three selected studies.

A. Heuristics organisation

As it is pointed out at the Methodology section, the heuristics are gathered in 6 groups according to the stages of the Buying Decision Process described by Engel et al. [5].

1) *Need Recognition and Problem Awareness* (Table II): *It is* the first step of the process. Without the recognition of the need a purchase cannot take place. In this stage, customers use search and navigation tools to find a product or service which covers their needs. The website has to stimulate the desire to purchase to enhance that needs.

2) *Information Search* (Table III): This is the next step customers may take after they have recognized their problems and needs in order to find out what they feel is the best solution. The customers search for information sources related to the products which can cover their needs.

3) *Purchase Decision Making* (Table IV): At this stage, consumers evaluate different products on the basis of varying product attributes to find the product which offers more benefits. After that process, the customers make a choice.

4) *Transaction* (Table V): The checkout is the last step in the Buying Decision Process. It has to be clear and trustworthy to drive the customers to finish a purchase.

5) *Post-Sales Behaviour* (Table VI): At this stage, companies should carefully create positive post-purchase communication to engage the customers. Also, the system has to facilitate tools to modify or follow customer's orders.

6) *Factors That Affect The UX During The Whole Purchase Process* (Table VII): There are many factors that affect whole purchase process and contribute in providing a satisfactory UX. One of the most important is trustworthiness because the customers have to believe that the website is reliable to finish a purchase.

B. Heuristics presentation

The heuristics are presented in the following six tables. It is shown according to the column order: the heuristics formulated as a questions to contrast with the website, a short explanation or examples to assist the rating decision making, the sources where the recommendation come from and the most appropriate rating system. It is recommended to add an additional column where experts can write their observations and reasons for scoring the heuristics in a manner of their choosing.

TABLE II. Need Recognition And Problem Awareness

Heuristics	**Description**	**Source(s)**	**Importance of factors**
	Search and navigation tools		
Is the navigation obvious enough throughout the related sections?	Customer should be able to move easily through the different sections. For instance, between the shopping cart, the detailed description of the products and the shipping information.	M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Does the website use a clear user-logical hierarchy of categories to classify products and to find them?	The categories should be easy to identify and differentiate by the customer.	T. P. Liang and H. J. Lai [8]; N. J. Lightner [9]; M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Do Category Pages include appropriate filters or facets by features?	They have to correspond customers' needs and be easy to undo. They are especially useful in shops that have a large amount of products.	N. J. Lightner [9]; T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)

Heuristics	Description	Source(s)	Importance of factors
Does the website provides a search box to locate products and information?	It must be visible at the top right of the page and it must continue throughout the whole site.	N. J. Lightner [9]; T. P. Liang and H. J. Lai [8]; M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Does the search have advanced features that allow for a limit to a great variety of criteria (features, categories, etc.)?	The advanced features have to correspond customers' needs This helps to retrieve the most relevant results.	N. J. Lightner [9]; T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Does the search engine provide the customer's expected results?	An analysis of customer searches must be made as a means to optimize search results.	T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Are there appropriate mechanisms, such as filters or facets to refine the search results?	After doing a simple or advanced search the results can be refined by applying these mechanisms. They have to correspond customers' needs and be easy to undo.	N. J. Lightner [9]; T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Do the pages and sub-pages provide orientation elements?	As a means to orientate it is necessary to use breadcrumbs, titles and subtitles.	T. P. Liang and H. J. Lai [8]; M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Does the checkout process includes a progress indicator at the top of the checkout pages?	Usually, it is a progress bar which indicates the steps that are missing to complete the purchase and the steps that have already been completed.	T. P. Liang and H. J. Lai [8]; M. Cao, Q. Zhang, and J. Seydel [3]	(Yes / No)
Does the website clearly display the "call to action buttons"?	This means that the buttons like "Add to Cart" or "Buy now" are easy to see and click. They should be located away from other buttons to avoid being clicked by mistake.	M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Does the website includes a site map?	This gives the user an overview of the site's areas in a single glance.	M. Cao, Q. Zhang, and J. Seydel [3]	(Yes / No)
Stimulating the desire to purchase			
Does the website use elements to draw customer's attention?	It can use banners, sounds and animations to focus customer attention on certain items or events.	T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Are the new products or special offers prominently advertised?	For instance, adding next to them "New!" or "On Sale."	T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Does the website shows the number of current visitors?	This is an index of the current shopping atmosphere that gives the customers a sense of store popularity.	T. P. Liang and H. J. Lai [8]	(Yes / No)

TABLE III. INFORMATION SEARCH

Heuristics	Description	Source(s)	Importance of factors
Is the information about the products accurate, informative and convincing?	It must include specifications and features with a non-technical and persuasive vocabulary.	M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Does the website provides value-added information and services?	For instance, lists of best sellers, compiled trends, link news related to the products and services, etc.	T. P. Liang and H. J. Lai [8]; M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Is the website content regularly updated?	Out of date content leads to customers distrusting the website.	M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Is the content based on the users' needs?	The content should be based on the customers needs instead of being based around the product description.	M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Are there multimedia resources to explain the products?	The site can use videos, images, audio and animations. These can involve long download times, thus it is necessary to strike a balance and optimize their use.	M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Is there any indicator about the product availability as soon as possible in the purchasing process?	If a product isn't available, it has to inform the customer when the product(s) will be available again.	N. J. Lightner [9]	(Yes / No)

c. Information Search Heuristics

TABLE IV. PURCHASE DECISION MAKING

Heuristics	Description	Source(s)	Importance of factors
Alternative Evaluation			
Is there enough information that relates to products or services?	It must provide all necessary information required to help the customer to make a decision. This include detailed descriptions of the features. For example, in the case of a book this would be the title, author, publisher, format, description and images. It should specify the price, availability of the product, the shipping conditions and customer reviews.	N. J. Lightner [9]; M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Are there product-related ratings and reviews?	It is not wise to build an own rating and review system because it will be seen as less trustworthy. In such a system, it is better to let negative reviews stand unless they are obscene or violate the law.	N. J. Lightner [9]; T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Is there a mechanism for the customer to indicate the usefulness of other customers reviews?	It can use a question, such as "Have you found this useful?" and as an answer "Yes" or "No". This displays how many people have found each review useful, and how many have not.	N. J. Lightner [9]	(1 2 3 4 5)
Are customers allowed to comment in other customers reviews?	The website should facilitate interaction between customers as they remain loyal to the site.	N. J. Lightner [9]	(Yes / No)
Does the website include product reviews published by the media?	It should add links to these reviews and the logos of the media to increase the trustworthiness of the website.	N. J. Lightner [9]; T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Are there tools to ease the comparison between different products?	For instance, it can assist the customer by providing a summary of the most important features - according to customers' needs - and costs of each product within a category. Alternatively, it can provide a tool to add products to a comparative table.	N. J. Lightner [9]; T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Are the order charges, such as taxes and shipping costs specified as soon as possible in the purchasing process?	If this information is hidden it causes distrust and shopping cart abandonment.	N. J. Lightner [9]	(1 2 3 4 5)
Is there information about the delivery dates?	Customers would like to know when the order is going to be delivered.	N. J. Lightner [9]	(Yes / No)
Does the website provide recommended products?	This can assist in finding products which might interest the customer, because they are similar to the products he purchased before or to products in which they expressed an interest. The related products can be inserted at the bottom of the product page, shopping cart or both.	N. J. Lightner [9]; T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Does the website provides products related to the selected product?	It should show accessories for the product which customer may want to purchase. This section can include between 2 and 6 related products and it is normally located at the product page or the shopping cart.	N. J. Lightner [9]	(1 2 3 4 5)
Choice			
Does the website has a shopping cart which is accessible from all the pages?	It must contain everything selected by the customer and the content should be accessible at anytime.	N. J. Lightner [9]; T. P. Liang and H. J. Lai [8]	(Yes / No)
Does the website incorporates a Wishlist?	It is a useful tool to manage products. For many users adding products to a wishlist is less committed than to a shopping cart.	N. J. Lightner [9]	(Yes / No)
Are there Intelligent Agents that can assist the customer?	Their primary role is as a shopping assistant, helping customers to solve problems. They can also be used to promote products.	T. P. Liang and H. J. Lai [8]	(Yes / No)

d. Purchase Decision Making Heuristics

TABLE V. TRANSACTION

Heuristics	Description	Source(s)	Importance of factors
Is there enough information to assist in the purchase process?	It could be useful for novice users to have section with a brief explanation. It also can be achieved by adding contextual help during the purchase process this can be achieved by introducing examples in the form fields.	N. J. Lightner [9]	(1 2 3 4 5)
Is the checkout process divided into logical steps?	An example of common logical steps would be "Shipping address and payment", "Shipping and gift options" (which include options like gift wrapping or adding a message), "Payment", "Order Summary" and "Order Confirmation".	N. J. Lightner [9]	(1 2 3 4 5)
If registration is required, is the process short and simple and does it demand only essential information?	If the customers consider that the information required by a field is not mandatory the chance of them falsifying data to speed themselves through the process increases. For this reason, it is useful to keep the process as short as possible and in the case of required information briefly explain why it is required.	T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Are there enough alternatives for the delivery of the order?	The customer should be able to choose the company they prefer, such as USPS, FedEx, UPS, DHS, etc. and consequently manage the speed of the delivery.	N. J. Lightner [9]; T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Does the website allow for enough payment options?	To avoid losing orders it should offer as many options as possible, such as Visa, MasterCard, bank account, electronic checks, PayPal or promotional codes. It is wise to display the logos of the payment options to make them more visible and trustworthy.	N. J. Lightner [9]; T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Are the different costs and discounts applied in the order and detailed before it is approved?	There has to be an order summary which specify the prices for each item, number of items, discounts applied, taxes and the total cost.	N. J. Lightner[9]	(1 2 3 4 5)
Is the button to confirm the purchase clearly visible?	It is usually labeled as "Buy" and it should be large, highlighted by color and linked to the order confirmation page.	N. J. Lightner [9]	(1 2 3 4 5)
Does the website provide different means for completing the order?	The site can facilitate the completion of the purchase by fax, phone, etc. Many online orders are not completed because of perceived safety.	N. J. Lightner [9]	(Yes / No)
Does the website show security logos?	It should show them specially in the checkout process to build trust. The logos may be related to the shipment (FedEx, UPS, etc.), payment options (Visa, PayPal, etc..) or security, like SSL.	T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Does the website inform of the level of security when paying by credit card?	It should demonstrate that it is a safe website to buy from.	N. J. Lightner [9]	(Yes / No)

e. Transaction Heuristics

TABLE I. TABLE VI. POST-SALES BEHAVIOUR

Heuristics	Description	Source(s)	Importance of factors
Does the system send a confirmation email after the customer's order?	The email should summarize the order and thank the customer. This generates a positive opinion from customer service.	N. J. Lightner [9]	(Yes / No)
Is it possible to track the status of an order from the customer account?	This section should allow the customer to consult a previous order and to track current orders.	N. J. Lightner [9]; T. P. Liang and H. J. Lai [8]	(Yes / No)
Can the customers manage their order(s) from the customer account?	The user should be able to modify or cancel orders. If returns are possible it should allow the tracking of its status.	N. J. Lightner [9]	(1 2 3 4 5)
Does the website allow the customer to return an item?	This boosts loyalty and potential purchases.	T. P. Liang and H. J. Lai [8]	(Yes / No)

f. Post-Sales Services Heuristics

TABLE VII. FACTORS THAT AFFECT THE USER EXPERIENCE DURING THE WHOLE PURCHASE PROCESS

Heuristics	Description	Source(s)	Importance of factors
Is the response time of the website reasonable?	Customers do not tolerate long waiting times. If the time to download a page is not reasonable for them they may leave the site.	M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Is the waiting time for the search results reasonable?	This depends on the size of the database. If the waiting time is going to be long, it would be wise to include small illustrations and animations to keep the customers waiting.	M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Is the interface's style consistent?	It reduces the cognitive load and facilitates the ease of orientation.	T. P. Liang and H. J. Lai [8]	(1 2 3 4 5)
Does the website presents an innovative and attractive image?	This helps to differentiate it from the competition and it is highly appreciated by customers. For this reason, they spend a longer amount of time on the site.	M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Is the website exciting?	It can include resources, such as interactive materials, downloadable applications, games, personalized information, etc. This benefits the playfulness and entertainment on the site and consequently the customer's satisfaction.	M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Does the website personalize any type of contact with the customer?	For example, when the system send an email it can include the customer's name.	M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Does the website offer the possibility for the customer to become a VIP?	This exclusivity increases the satisfaction and loyalty of the customers who spend more money.	T. P. Liang and H. J. Lai [8]	(Yes / No)
	Trust building		
If personal information is required by the website, does it have the Privacy Policy available?	It should include all possible uses of their personal information. The footer is the standard location to link this information. It has to be easily understandable by the customers.	N. J. Lightner [9]; M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Does it have the shipping, return or exchange Policy and other shop rules available?	This information can be included in the FAQs or in a specific section of the store. These resources have to be easily understandable by the customers.	N. J. Lightner [9]	(1 2 3 4 5)
Does the website has safety certificates granted by external companies?	It is important to get them and let customers know about them. These are companies like VeriSign or ControlScan. It must use the secure SSL technology as well.	N. J. Lightner [9]; T. P. Liang and H. J. Lai [8]	(Yes / No)
Is contact information visible during the purchase process?	The telephone number should be shown at least to answer any questions about orders.	T. P. Liang and H. J. Lai [8]	(Yes / No)
Does the website provide different means for customers to contact the company?	For example forms, email and phone as these improve customer confidence.	N. J. Lightner [9]; M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Is the best way to contact the company clarified for each type of concern?	It should specify the best means of contact for each type of question. For example, if it's a technical question, a matter of sales or about returning an item.	N. J. Lightner [9]	(1 2 3 4 5)
Does the website gives the address of the company?	It can be shown in the information about the company, in the FAQs and in the checkout process to give credibility to the page.	N. J. Lightner [9]	(Yes / No)
Does the website has a FAQ section that covers common customer questions?	It has to cover the most common questions detected by the customer service. For example, it can answer questions concerning the return policy. The information included has to be easily understandable and useful.	N. J. Lightner [9]	(1 2 3 4 5)
Does the company responds to comments and concerns expressed by the customers?	Either through private means like email or public, like a product review.	N. J. Lightner [9]; M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)
Is the appearance of the website safe and reliable?	Security is independent from the graphic design but the perception of security is very important for the customer. For instance, the introduction of elements, such as the image of a padlock affects positively this perception.	M. Cao, Q. Zhang, and J. Seydel [3]	(1 2 3 4 5)

g. Heuristics that affect the UX during the whole purchase process

 ISBN: 978-1-61208-325-4

IV. Conclusion and Future Work

As we pointed out in the methodology section, the heuristics presented above are derived from functional requirements expressed by customers and from factors that capture the quality of e-commerce websites. Therefore, they do not equally cover every UX's facet mentioned in the introduction. The only aspect that any of the three studies have considered is the accessibility. None the less, it is quite common to check it separately with online tools that automatically analyze it and provide quantitative results [4]. Anyway, as a future work, it is planned to integrate accessibility requirements in the proposed model to improve the consistency of the evaluation.

The presented heuristic set is an assisting tool for evaluators based on proved standards to review the aspects of a website that can affect UX. It is an open list that is going to be extended in a broader study by adding more recommendations to create a greater and more exhaustive set of heuristics. The set will be updated and adapted by modifications, updates and additions because of the continuous evolution of e-commerce sites. The follow up study's goal is to cover as many B2C websites and UX's aspects as possible. In this way, the set can serve a wide range of e-commerce sites that may require only small adjustments depending on the website features. Thus, the set might require some revision to discard the sections and heuristics which do not apply.

Moreover, it is necessary to add that the current set of heuristics has some more limitations apart from the ones mentioned. The heuristics understandability has been checked by two experts, but the set has not yet been used to review any e-commerce websites. Consequently, it is not possible to prove empirically its efficiency.

Finally, as future work, it has been planned to follow this research introducing the presented heuristics into the Open-HEREDEUX repository [12][13]. This will enable to start an experimental test with two objectives: enhance the database of Open-HEREDEUX heuristics (introducing a specific set for e-commerce systems) and validate the set of guidelines presented in this paper. The second goal will consist in evaluating different e-commerce sites (between 5 and 10) in order to obtain and analyze the results. This also will serve to refine the e-commerce set of heuristics.

References

[1] Baymard Insitute, “18 Cart abandonment rate statistics” [Retrieved: January, 2014]. Available at URL:
<http://baymard.com/lists/cart-abandonment-rate>

[2] S. Bauer, “It works for you: A user-centric guideline to product pages”, in Smashing Magazine, January 10, 2012 [Retrieved: January, 2014]. Available at URL:
<http://uxdesign.smashingmagazine.com/2012/01/10/it-works-for-you-a-user-centric-guideline-to-product-pages/>

[3] M. Cao, Q. Zhang, and J. Seydel, “B2C E-commerce web site quality: An empirical examination”, in Industrial Management & Data Systems, 105 (5), pp. 645-661.

[4] Web Accessibility Initiative, “Complete list of web accessibility evaluation tools” [Retrieved: January, 2014]. Available at URL:
<http://www.w3.org/WAI/RC/tools/complete>

[5] J. F. Engel, R. D. Blackwell, and D. T. Kollat, Consumer behavior, 3rd ed. Hinsdale: Dryden Press, 1978.

[6] C. Holst, “Fundamental guidelines of e-commerce checkout design”, in Smashing Magazine, April 6, 2011 [Retrieved: January, 2014]. Available at URL:
<http://uxdesign.smashingmagazine.com/2011/04/06/fundamental-guidelines-of-e-commerce-checkout-design>

[7] ISO DIS 9241-210:2008, “Ergonomics of human system interaction – Part 210: Human-centered design for interactive systems (formerly known as 13407)”. Switzerland: International Standardization Organization, 2008.

[8] T. P. Liang, and H. J. Lai, “Effect of store design on consumer purchases: An empirical study of on-line bookstores”, in Information & Management, 39 (6), pp. 431-444.

[9] N. J. Lightner, “Evaluating e-commerce functionality with a focus on customer service”, in Communications of the ACM, 47 (1), pp. 88-92.

[10] L. Masip, [et al.], “A design process for exhibition design choices and trade-offs in (potentially) conflicting user interface guidelines”, in Lecture Notes in Computer Science, Volume 7623, pp 53-71

[11] L. Masip, T. Granollers, and M. Oliva, “A Heuristic Evaluation Experiment To Validate The New Set Of Usability Heuristics”, in Proceedings of the 8th International Conference on Information Technology: New Generations, (Washington, DC, USA, 2011), IEEE Computer Society, pp. 429-434.

[12] L. Masip, M. Oliva, and T. Granollers, “OPEN-HEREDEUX: open heuristic resource for designing and evaluating user experience”, in Proceedings of the 13th IFIP TC 13 International conference on Human-Computer Interaction (INTERACT'11), Pedro Campos, Nuno Nunes, Nicholas Graham, Joaquim Jorge and Philippe Palanque (Eds.), Part IV. Springer-Verlag, Berlin, Heidelberg, pp. 418-421.

[13] L. Masip, M. Oliva, and T. Granollers, “The open repository of heuristics”, in Proceedings of the 13th International Conference on Interacción Persona-Ordenador (INTERACCION '12). ACM, New York, NY, USA.

[14] P. Morville, “User Experience Design”, in Semantic Studios, June 21, 2004 [Retrieved: January, 2014]. Available at URL:
<http://semanticstudios.com/publications/semantics/000029.php>

[15] J. Nielsen, “Heuristic evaluation”, in Usability Inspection Methods, J. Nielsen and R.L. Mack, Eds. New York: John Wiley & Sons, 1994, pp. 25-62.

[16] J. Vanderdonckt, “Development milestones towards a tool for working with guidelines”, in Interacting with computers, 12 (2), pp. 81-118.

 ISBN: 978-1-61208-325-4

Sonification of Large Datasets in a 3D Immersive Environment: A Neuroscience Case Study

Panagiota Papachristodoulou, Alberto Betella
SPECS, NRAS, Universitat Pompeu Fabra
Barcelona, Spain

Paul F.M.J. Verschure
SPECS, NRAS, Universitat Pompeu Fabra and ICREA, Institució Catalana de Recerca i Estudis Avançats
Barcelona, Spain
Email: paul.verschure@upf.edu

***Abstract*—Auditory display techniques can play a key role in the understanding of hidden patterns in large datasets. In this study, we investigated the role of sonification applied to an immersive 3D visualization of a complex network dataset. As a test case, we used a 3D interactive visualization of the so called, connectome of the human brain, in the immersive space called "eXperience Induction Machine (XIM)". We conducted an empirical validation where subjects were asked to perform a navigation task through the network and were subsequently tested for their understanding of the dataset. Our results showed that sonification provides a further layer of understanding of the dynamics of the network by enhancing the subjects' structural understanding of the data space.**

Keywords-sonification; XIM; networks; neuroscience; complex data; auditory display.

I. Introduction

Visual representations of data (also called visual displays) have a long successful history in scientific research. They have been employed widely and commonly in most of the traditional HCIs [1] playing a key role in uncovering hidden structures and meaning in massive collections of data. However, relying solely on visual representations of information can present some limitations when dealing with large amounts of multidimensional data [2]. In some cases, visual representation of data can also lead to cognitive saturation [3].

Humans have the capability to detect subtle temporal patterns in sounds [4]. Due to the auditory system's integrative properties, sound can be efficiently used to enhance visual representations without creating information overload for the user [5]. In addition, sound might be more appropriate to convey dynamical information as opposed to vision [6].

For this reason, in the last few decades, scientists have been introducing auditory cues as a means to convey information, giving rise to a new field called auditory display. As a subset of this, the field of sonification has emerged with the purpose of providing a new tool for the comprehension of complex data and generate new insights. A number of sonification models and techniques have been developed [7], [8], [9] with the purpose of converting data into meaningful information.

Many studies have investigated the role of sound for the representation of complex datasets [10], [11], [12], [13]. However, most of these studies focused on the effectiveness of visual and auditory displays separately, yet a few among them have investigated the role of sound as enhancer of a graphical visualization (i.e., multimodal display) [14], [15], [16], [17].

Here, we investigated how sonification may function when integrated in a multimodal display of large neuroscience datasets such as the connectome [18]. We previously developed the so called XIM Neuroscience application[19], a 3D interactive framework to visually represent and explore the massive connectivity of the human connectome, which is "a comprehensive structural description of the network of elements and connections forming the human brain"[18]. For the purpose of our experiment, we displayed the Neuroscience application in XIM, an immersive room equipped with a number of sensors and effectors that have been constructed to conduct experiments in mixed reality [20].

Our hypothesis was that the introduction of sound in a multimodal display could enhance the understanding of complex data describing the brain.

We paired the 3D visualization of a connectome dataset [21] in XIM with a sonification system intended to represent the changes in the dynamics related to different regions of the network, thus providing the user with an extra channel of information. For the empirical evaluation, we assessed whether sound could augment the understanding of the network's dynamics.

This paper is composed of three different sections. Section II introduces the methods and materials along with the sonification model that was adopted. Section III presents the results of the experiments and section IV provides an overview of the conclusions.

II. Methods

A. The "eXperience Induction Machine (XIM)"

The XIM [20] is a multi-user mixed-reality space covering a surface area of 5.5 x 5.5m equipped with a number of sensors and effectors. XIM effectors include computer graphics content projected via 8 projectors on 4 separate walls, a luminous

 ISBN: 978-1-61208-325-4

interactive floor, movable lights and sonification system. For the purposes of this project, four projectors were used as a visual display of the connectome network on 4 separate walls and two speakers in the left and right corners of the room to project the auditory display.

B. The Neuroscience Application

The Neuroscience application is a system that was previously built in XIM [19] that uses multimodal input and output and permits the embodied exploration of connectome datasets. In this study, we adopted the connectome dataset from Hagmann et al. [21] (data source http://www.cmtk.org/datasets/homo_sapiens_01.cff). The dataset is composed of 998 regions of interest (ROIs), 28000 unidirectional connections and 66 anatomical subregions in the left and right hemisphere of the human brain.

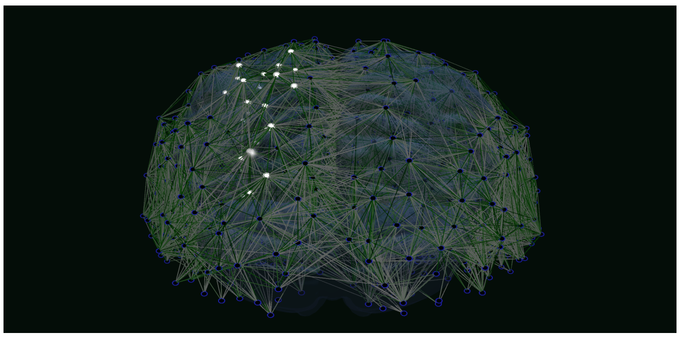


Figure 1. 3D visualization of the connectome in XIM.

The connectome network is visually represented in XIM. Edges are represented as tubes mapped to different shades of green in accordance to their strength. Nodes are represented as spheres and are highlighted when the region they belong to is selected by the user. The network's anatomical regions contain the nodes and are spatially organized in accordance with a standard brain atlas (Talairach coordinates of ROIs[22]) (Fig. 1). The system is implemented using Unity 3D (http://unity3d.com).

C. Sonification methods

The goal of our sonification was to acoustically represent changes in the connectome network's parameters while navigating through its different structures.

First, we analyzed the distribution of the values of the parameters in the network in terms of a) number of nodes, b) number of connections, c) average strength of each region and d) brain hemisphere.

Second, two distinct sound sources were chosen whose parameters were mapped to the network's parameters: a) a short sound sample (grain sound of 16 ms), acoustically perceived as a "click" and b) a soft multi-layered ambient sound, which could be characterized as a drone sound. For each one of the 66 brain regions, the sound parameters used for the sonification were: a) the repetition rate (playback frequency) of the grain sound, which was mapped to the number of nodes, b) the pitch of the ambient sound, mapped to the number of connections, c) the loudness of the grain sound, representing the average strength of each region and d) panning, to identify the brain hemisphere (Table I).

TABLE I. MAPPING BETWEEN THE NETWORK'S CHARACTERISTICS AND THE CORRESPONDING SOUND PARAMETERS.

Connectome Characteristics	Sound source	Sound Parameter
Number of connections	Sound grain of 16 ms	Rate of repetition
Number of nodes	Ambient sound at 380 Hz enhanced by a pure sine wave at the same frequency	Pitch
Average Strength	Sound grain of 16 ms	Loudness

The sonification engine was built using Pure Data (http://puredata.info/) and the communication between Unity and Pure Data was implemented using OpenSoundControl (OSC) messages.

1) Nodes: A narrow-band short sound sample with the duration of 16ms was chosen for the sonification of the nodes. The idea was to take inspiration from granular synthesis (a method that operates on the microsound time scale [23]) and simulate laboratory sound recordings of neural activity. These microsound characteristics allowed the auditory display of the micro-structures of the brain. By doing so, we associated the ROIs in the brain (i.e., the nodes of the network) to a sound texture composed of a large number of sound grains. The number of nodes determined the time interval between the grains, which defined the number of clicks that would be played back by our sonification engine.

The number of nodes was mapped to an interval that determined the metro (or rate) of the playback in an inversely proportional relation (i.e., lower number of nodes resulted in larger time intervals between the grains and vice versa). In order to avoid a mechanical and isochronous repetition of the grain sound, the output in milliseconds was also used as input for a random function. Likewise, a low number of nodes resulted in large range of intervals of the random function. The output of this function was used as a delay time, which was added to the final playback repetition.

2) Connections: Previous literature validated the role of pitch in the sonification of complex data for the purpose of different tasks [11], [12]. In addition, pitch offers large resolution since the human auditory system is capable of detecting very subtle changes in a very large range of frequencies [24]. For this reason, we mapped the pitch to the number of connections of each connectome region. A sound parameter with a large resolution, such as the pitch allowed for the auditory presentation of a wide range of number of connections.

We adopted a soft ambient sound sample to provide a more pleasant auditory experience to the user than a pure sine wave [25]. First, a spectral analysis was conducted with SonicVisualiser (http://www.sonicvisualiser.org), which revealed that the most prominent frequency of the sample was at 380Hz. This value was adopted to define the upper limit of the frequency range. Hence, the number of connections was mapped to a frequency range spanning from 60Hz to 1200Hz. Frequencies above 1200Hz would have resulted in an unpleasant and disturbing experience for the user.

To accomplish the sound mapping in our sonification engine, we used a pitch shift object in pure data. This object shifts the frequencies of the sound resulting in lower and higher final pitch. Three groups were defined for the whole range of numbers of connections. Each group was mapped separately into a subset of values using linear mappings, which depended on the distribution of the connections values in the network. These values were used to transpose the sound through the pitch shift function resulting in the final fundamental frequency. Finally, the amplitude of the sound was adjusted according to the equal - loudness contours [24]. This correction ensured that all the frequencies would present the same level of loudness.

3) Strength: Loudness is a perceptual sound parameter that has been used previously in sonification models [12]. For this reason, we used loudness to represent the strength of the connections in the network. Since the manipulation of the amplitude and the frequency of the same sound source simultaneously could lead to unperceived and confusing changes [26], we chose to modulate the loudness of grain sound used for the nodes, as opposed to the ambient sound of the connections.

We calculated the average strength for each region, which was mapped to the amplitude of the grain sound. According to the literature, "the most effective use of loudness change usually occurs when changes in loudness are constrained to two or three discrete levels that are mapped to two or three discrete states of the data being sonified. In this way, discrete changes in loudness can be used to identify categorical changes in the state of a variable or to indicate when a variable has reached some criterion value" [27]. For this reason, 3 groups of average strength were defined for low, medium and high levels of loudness so that the user would easily perceive the changes from one group to the other.

We adopted a linear mapping for the three groups. We defined a lower and upper limit for each group and implemented a scaling of the values of average strength into root mean square (rms) values of the sound wave. The rms values were then multiplied by the signal in order to determine the value of the amplitude of the grain sound. Additionally, a reverb was added to the higher values to enhance the perception of higher strength for the correspondent regions of the network. The scaling factor that was used for the mapping in rms depended on the distribution of the values in the network. Thus, for the first two groups the variation was smoother and for the third group it was more rapid.

4) Brain Hemispheres: For the brain hemispheres we used sound panning. This allowed for the discrimination between left and right hemisphere since the direction of the sound source makes the understanding of the location of the regions in the network more intuitive [24]. Binary signals were received from Unity, which were then mapped to the two speakers through the pan object in pure data.

D. Empirical Evaluation

The aim of this study was to assess whether sound could enhance the understanding of the network's characteristics displayed in XIM and their properties during a navigation task through the network.

Our hypothesis was that sonification would enhance the accuracy in the estimation of the characteristics of the connectome, thus providing a higher understanding of its dynamical changes and a more precise discrimination between the brain regions. By implementing a multimodal experiment using sonification, based on two contrasting sound textures (i.e., sound grains and a long drone), we aimed to empirically validate how sound could be used as an integrative tool to enhance the human ability to detect structural aspects of the connectome.

25 healthy adults (15 females, mean age=29.53, SD $\pm$5.6) with normal or corrected vision and hearing were recruited. The subjects had no prior knowledge of the connectome dataset that they were exposed to during the experiment.

To avoid between-subjects differences in auditory recognition skills, the experiment followed a paired-samples design, where each subject was exposed (in random order) to two conditions: a) only visualization and b) both visualization and sonification. The independent variable was the presence of sound (i.e., sonification condition) versus the absence of sound (visualization condition), while the dependent variable was the quantitative estimation of the properties related to different regions in the network during a navigation task.

E. Experimental setup

The subjects' task was to estimate the number of nodes, connections and the average strength of each brain region and to understand the changes of these values during the navigation. They were presented with different brain regions and were asked to mark their answers in closed-ended questions. The experiment consisted of a training session, followed by two experimental sessions.

The XIM was setup and equipped to provide an immersive experience to the subjects. The lights were dimmed, a table and a comfortable chair were placed in the middle of the room at a symmetrical distance from the speakers to allow a balanced auditory experience for all the subjects. The navigation was done with a standard keyborad and mouse.

1) Regions selection: We preselected a number of brain regions to be presented during the training and the first session. 14 regions of the network were chosen for the training session, resulting in 7 trials for each condition. The subjects were exposed to regions with similar properties (number of nodes, connections and average strength) in both conditions.

For the first experimental session, we selected 36 regions in total for both conditions resulting in 18 trials (9 pairs of regions) for each condition. For each one of the 3 parameters (nodes, connections, average strength) at least one pair of regions presented evident acoustical difference in the parameters (because of the high difference in their values).

2) Training Session: Visual information displays (e.g. graphs) are often familiar to the users given their former education (they are taught since young age) and informal experience [28]. However, complex auditory displays can be not as intuitive as visual representations. For this reason, most of the sonification studies indicate the importance of the training sessions [29], where the subjects can get acquainted with the network's characteristics and their corresponding visual and auditory meaning.

The task of the subjects during the training session consisted in the estimation of the values of the connectome's characteristics for the preselected regions. Specifically, the subjects were asked to estimate the number of nodes, connections and average strength of each region and in which hemisphere the region was located. The regions were presented in random order in both conditions. After the exposure to a region, subjects were asked to mark their answers in a closed-ended questionnaire presenting different value ranges in accordance with the properties of each area and their distribution in the network. Subsequently, subjects were given feedback on their answers so that they could learn their possible mistakes.

3) Session 1: In the first experimental session, the subjects were exposed to pairs of different regions (as specified in Section II-E1). First, they were informed about the characteristic they had to evaluate (number of nodes, connections or average strength). Then, they were asked to choose whether the second region presented had a higher or lower value of the parameter measured compared to the first region. After the presentation of each pair, they were asked to mark with plus or minus their answer in a questionnaire. Subjects where allowed to ask for a repetition of the trial in case they wanted to listen (and see) one more time the pair of regions.

4) Session 2: In the second experimental session, subjects were asked to navigate freely through the connectome. Their task consisted in finding a region with certain values (e.g., a region with low number of connections between 200-400 connections, a region with medium-high number of nodes between 20-30 nodes, etc.) as asked by the experimenter at the beginning of each trial. There were no time constraints and subjects were given a printed reference table showing the minimum and maximum values for each one of the parameters. After each answer, subjects were given feedback and proceeded to the following question. The task was repeated twice, once for each condition. Each subject completed 6 trials for each condition, resulting in a total of 12 trials.

5) Score attribution: We defined the score attribution criteria for the two experimental sessions. For each subject, we calculated a score for each one of the parameters, along with an overall score per condition. Table II summarizes the tasks and the score attribution criteria for each one of the sessions.

TABLE II. TASK FOR EACH SESSION AND SCORE ATTRIBUTION CRITERIA.

Session	**Task**	**Score Attribution Criteria**
Session 1	Estimation of higher and lower values between pairs of regions	Correct answers were attributed with 1, otherwise with 0
Session 2	Finding regions with certain characteristics in predetermined ranges of values	Error deviation from correct answer in predetermined scale designed ad hoc for the evaluation

For the first session, the score was based on the number of correct and wrong answers. A score of 1 was assigned to the questions answered correctly and 0 to wrong answers. A total score from 0 to 9 was calculated for each of the two conditions for each subject. In the second session, we measured the absolute distance from the correct answer [30] and divided the resulting value by the number of possible answers, thus obtaining normalized scores between 0 and 1, representing the maximum and the minimum score respectively. We computed these values according to the following formula:

$$\beta_i = 1 - \frac{|i-k|}{N} \tag{1}$$

where β is the normalized score, i the subject's answer, k the correct answer and N the total number of probable answers.

III. RESULTS

During Session 1, we measured the subjects' accuracy in estimating the higher or lower values for the three parameters of the network (number of nodes, number of connections and average strength) between pairs of regions. The data collected in session 1 was submitted to a Wilcoxon test between the two conditions (visualization and sonification). The correct answers for both conditions were calculated. The sonification condition obtained a significantly higher score (z=-2.96, $p < .05$) as opposed to the visualization condition (Fig. 2a). The means of the correct answers of the subjects for each one of the parameters of the network are presented in Table III.

Wilcoxon tests were conducted for each one of the parameters between the two conditions. Although not statistically significant, the means for the sonification condition were higher than their counterparts in the visualization condition (Table III).

Since the experiment followed a paired samples design, the data were tested for order effect. No order effect was found.

TABLE III. MEANS OF THE SCORES FOR THE CORRECT ANSWERS BETWEEN THE TWO CONDITIONS.

	Session 2 Estimation of higher or lower values			
	Visualization		Sonification	
	Mean	Standard Deviation (SD)	Mean	Standard Deviation (SD)
Nodes	2.32	0.80	2.41	0.60
Connections	2.12	0.97	2.48	0.82
Average Strength	1.76	0.66	2.16	0.99
Total	6.00	1.52	7.00*	1.57

We conducted a Spearman's correlation test between the musical background of the subjects and the their estimations for higher and lower values between pairs of regions in the sonification condition. The results revealed a significant negative correlation between the correct answers of the subjects and their musical background (ρ=-0.37, $p < .05$). Subjects with higher formal musical training scored lower than subjects who didn't have a musical background. These results explain the skewness towards the lower score of the score distribution in the sonification condition (Fig. 2a). As a follow up, we conducted a further Wilcoxon test by excluding from the analysis the scores of the subjects with higher formal music training, seven subjects were excluded. We obtained significantly higher scores for the sonification condition as compared to the visualization condition (z=-2.44, $p < .05$) and a normal distribution in the scores (Fig. 2b).

During Session 2, we measured the accuracy of the subjects to find a region that would present specific values of each one of the three parameters (number of nodes, connections and average strength). For this experimental session, a dependent

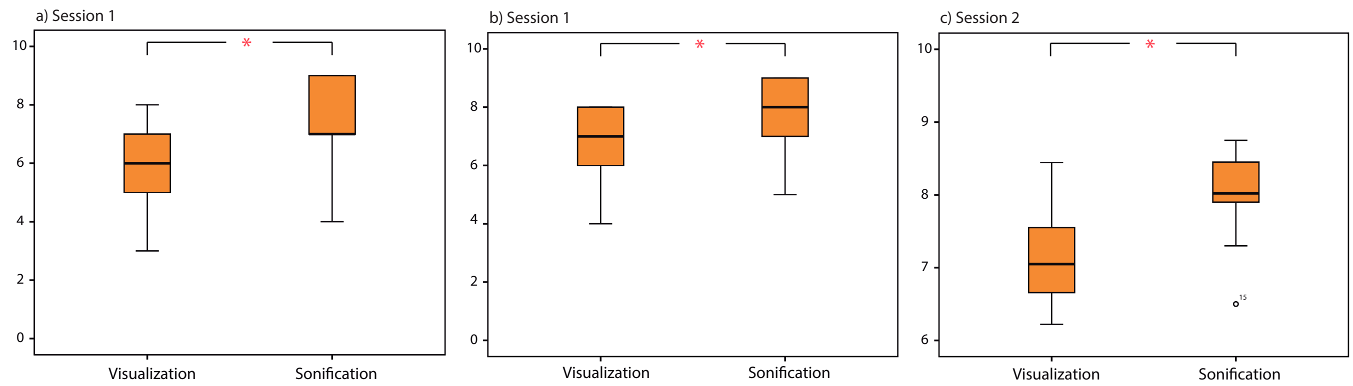


Figure 2. a) Boxplot showing the significantly higher score ($p < .05$) obtained for the sonification condition in Session 1. b) Boxplot showing the reduction of skewness in the data for Session 1, after the exclusion of subjects with high level of musical background. c) Boxplot showing the significant score obtained for the sonification condition in Session 2 ($p < .001$).

T-test was conducted for the total scores obtained from the error deviation measurements. The normalized scores for each subject were calculated for each one of the regions successfully found by the subject through the navigation in the network. The average score was calculated for each subject in both conditions. A significantly higher score was obtained for the sonification condition (t(24)=-5.03, $p < .001$) when compared to the visualization condition (Fig. 2c).

Wilcoxon tests for the scores of the separate parameters were conducted. The scores obtained were higher for the sonification condition for all the parameters.

TABLE IV. MEANS OF THE SCORES OBTAINED FROM THE ERROR DEVIATION FOR THE CORRECT ANSWERS IN BOTH CONDITIONS. THE STATISTICALLY SIGNIFICANT RESULTS ARE INDICATED WITH AN ASTERISK.

	Session 3 Error deviation			
	Visualization		Sonification	
	Mean	Standard Deviation (SD)	Mean	Standard Deviation (SD)
Nodes Low	0.94	0.09	0.96	0.10
Nodes High	0.90	0.14	0.95	0.11
Connections Low	0.48	0.35	0.85*	0.21
Connections High	0.83	0.21	0.87	0.13
Strength Low	0.77	0.29	0.87	0.13
Strength High	0.83	0.16	0.84	0.14
Total	7.15	0.62	8.03*	0.54

In addition, we found a statistically significant difference in score between the two conditions for the regions with low number of connections. The sonification condition obtained a significantly higher score (z=-3.42, $p < .05$) compared to the visualization (Table IV).

IV. CONCLUSION

In this study, we investigated the effect of sonification on the visualization of a complex network dataset. As a test case, we used a 3D interactive visualization of the human brain connectome displayed in the mixed reality "eXperience Induction Machine (XIM)".

We added a dynamic sonification to the 3D visualization of the connectome dataset and we conducted an experiment, where subjects were asked to perform a navigation task through the network and subsequently were tested for their understanding of the connectome dataset.

The parameters of the connectome network (i.e., nodes, connections, average strength and brain hemisphere) were mapped to repetition rate, pitch, loudness and panning, respectively and the estimation of the values in the two conditions (absence and presence of sound) was measured.

The empirical validation consisted of a training session followed by two experimental sessions. After the training, in the first session we measured the accuracy in estimating the higher or lower values of the three parameters between pairs of regions in the network. In the second session, we measured the accuracy of the subjects in finding regions within given range of values in the network parameters. The sonification condition obtained significantly higher scores in both experimental sessions.

Furthermore, the analysis of the data collected in the second experimental session showed that subjects obtained higher scores for the regions with low number of connections in the sonification condition during the free navigation through the network. In complex networks, such as the human connectome, connections (or edges) can't be easily discriminated visually. Here, we have shown that the pitch sound parameter can act as an enhancer in the understanding of structural connections.

The analysis of the results collected during the second session revealed that subjects obtained lower scores for the strength in the sonification condition for specific regions. This may be due to the interaction of the sound parameters (repetition rate and loudness). When an area presented high number of nodes associated to low strength, the interactions of these two parameters could have confused the subjects and these regions may have been associated to higher values of average strength. Further research on the interaction of the sound parameters will provide better understanding of the auditory perception.

In addition, we found a negative correlation between the scores obtained in the sonification condition and the musical formal training of the subjects. In the literature, there is no agreement about the relation of the musical background in auditory tasks. One explanation of this apparently counterintuitive result, as suggested by Walker and Nees [31], may lie on the fact that the skills acquired while undertaking musical studies, especially during the childhood, could be removed after many years. Another explanation for this result could be due to the fact that the measurement of musical background through a single question does not account for an accurate assessment; a more specific test to measure musical abilities should be administered.

In conclusion, the presence of sound enhanced the performance of the subjects (in terms of structural understanding of a network), hence we retain our alternative hypothesis.

Further improvements will include the sonification of additional parameters of the networks (hubs, clusters, etc.) along with the use of audio descriptors and spectral modelling techniques to validate which features, among the sound parameters, are more effective in the understanding of a complex network dataset.

V. ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7-ICT-2009-5) under grant agreement n. 258749 [CEEDS]. The Generalitat de Catalunya (CUR; Departament d'Innovació, Universitats i Empresa) and the European Social Fund are supporting this research. Thanks to Sebastian Mealla, Enrique Martínez Bueno, Jonatas Manzolli, Xerxes D. Arsiwalla, Anna Mura.

REFERENCES

[1] B. J. Fry, "Computational Information Design," Sciences New York, vol. Ph. D, no. May 1997, 2004, p. 170.

[2] A. Eldridge, "Issues in auditory display." Artificial Life, vol. 12, no. 2, 2006, pp. 259–274.

[3] N. Bearman and P. F. Fisher, "Using sound to represent spatial data in ArcGIS," Computers & Geosciences, vol. 46, Sep. 2012, pp. 157–163.

[4] T. Hermann and H. Ritter, "Listen to your Data : Model-Based Sonification for Data Analysis," Advances in Intelligent Computing and Multimedia Systems, vol. 8, 1999, pp. 189–194.

[5] G. Kramer, B. N. Walker, T. Bonebright, P. Cook, J. H. Flowers, N. Miner, and J. Neuhoff, "Sonification report: Status of the field and research agenda," Quantum, 1999, p. 38.

[6] A. S. Bregman, Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press, 1990.

[7] T. Hermann, "Taxonomy and definitions for sonification and auditory display," in Proceedings of the 14th International Conference on Auditory Display, Paris, France, 2008, pp. 1–8.

[8] A. de Campo, "Toward a data sonification design space map," Proceedings of the 13th International Conference on Auditory Display, Montreal, Canada, 2007, pp. 342–347.

[9] T. Hermann and H. Ritter, "Sound and Meaning in Auditory Data Display," in Proceedings of the IEEE, vol. 92, no. 4, 2004, pp. 730–741.

[10] S. Pauletto and A. Hunt, "Interactive sonification of complex data," International Journal of Human-Computer Studies, vol. 67, no. 11, 2009, pp. 923–933.

[11] B. N. Walker, "Magnitude Estimation of Conceptual Data Dimensions for Use in Sonification," Journal of Experimental Psychology, vol. 8, no. 4, 2002, pp. 211–221.

[12] J. H. Flowers, "Thirteen Years of Reflection on Auditory Graphing : Promises , Pitfalls , and Potential New Directions," in Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland, 2005, pp. 1–5.

[13] T. Hermann, "Sonification for Exploratory Data Analysis," Ph.D. dissertation, Bielefeld University, Bielefeld,Germany, 2002.

[14] S. Bly, "Presenting information in sound," in ACM CHI '82 Conference on Human Factors in Computer Systems. ACM, New York, 1982, pp. 371–375.

[15] P. Janata, "Marketbuzz: Sonification of real-time financial data," Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia, 2004, pp. 4–10.

[16] S. C. Peres and D. M. Lane, "Auditory graphs: The effects of redundant dimensions and divided attention," in Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland, 2005, pp. 169–174.

[17] K. V. Nesbitt and S. Barrass, "Evaluation of a Multimodal Sonification and Visualisation of Depth of Market Stock Data," in Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan, 2002, pp. 1–5.

[18] O. Sporns, G. Tononi, and R. Kötter, "The human connectome: A structural description of the human brain." PLoS computational biology, vol. 1, no. 4, 2005, p. e42.

[19] A. Betella, E. Martínez, R. Zucca, X. D. Arsiwalla, P. Omedas, S. Wierenga, A. Mura, J. Wagner, F. Lingenfelser, E. André, D. Mazzei, A. Tognetti, A. Lanatà, D. De Rossi, and P. F. M. J. Verschure, "Advanced interfaces to stem the data deluge in mixed reality: placing human (un)consciousness in the loop," in ACM SIGGRAPH 2013 Posters, ser. SIGGRAPH '13. New York, NY, USA: ACM, 2013, pp. 68:1—-68:1.

[20] U. Bernardet, S. Bermudez i Badia, and P. F. Verschure, "The experience induction machine and its role in the research on presence," in Proceedings of the 10th Annual International Workshop on Presence, 2007, pp. 329–333.

[21] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, and O. Sporns, "Mapping the structural core of human cerebral cortex." PLoS biology, vol. 6, no. 7, 2008, p. e159.

[22] J. J. Talairach, P. Tournoux, and O. Missir, Referentially oriented cerebral MRI anatomy : an atlas of stereotaxic anatomical correlations for gray and white matter / Jean Talairach, Pierre Tournoux, with the collaboration of Odile Missir ; [translated from the French original manuscript by Baris Tu. Stuttgart ; New York : G. Thieme Verlag ; New York : Thieme Medical Publishers, 1993.

[23] C. Roads, Microsound. The MIT Press, 2004.

[24] S. Carlile, "Psychoacoustics," in The Sonification Handbook, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Logos Publishing House, 2011, ch. 3, pp. 41–61.

[25] L. M. Brown, S. Brewster, R. Ramloll, M. Burton, and B. Riedel, "Design guidelines for audio representation of graphs and tables," in Proceedings of the 9th International Conference on Auditory Display (ICAD2003), E. Brazil and B. Shinn-Cunningham, Eds., Boston University Publications Production Department. Boston University Publications Production Department, 2003, pp. 284–287.

[26] J. G. Neuhoff, G. Kramer, and J. Wayand, "Pitch and loudness interact in auditory displays: can the data get lost in the map?" Journal of experimental psychology. Applied, vol. 8, no. 1, 2002, pp. 17–25.

[27] J. G. Neuhoff, "Perception, Cognition and Action in Auditory Displays," in The Sonification Handbook, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Logos Publishing House, 2011, ch. 4, p. 6385.

[28] S. Ferguson and D. Cabrera, "Exploratory sound analysis: Sonifying data about sound," in Proceedings of the 14th International Conference on Auditory Display, Paris, France, 2008, pp. 1–8.

[29] T. L. Bonebright and J. H. Flowers, "Evaluation of Auditory Display," in The Sonification Handbook, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Logos Publishing House, 2011, ch. 6, pp. 111–144.

[30] B. Walker, J. Lindsay, J. Godfrey, and C. Street, "The Audio Abacus : Representing Numerical Values with Nonspeech Sound for the Visually Impaired," ACM SIGACCESS Accessibility and Computing, vol. 77-78, 2004, pp. 9–15.

[31] B. Walker and M. A. Nees, "Theory of Sonication," in The Sonification Handbook, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Logos Verlag, 2011, pp. 9–40.

Design Guidelines and Design Recommendations of Multi-Touch Interfaces for Elders

Bruno Loureiro[1], Rui Rodrigues[1,2]
[1]DEI/FEUP - Department of Informatics Engineering, Faculty of Engineering, University of Porto, Porto, Portugal
[2] INESC – TEC, Porto, Portugal
{bruno.loureiro, rui.rodrigues}@fe.up.pt

***Abstract*— The usage of multi-touch interfaces on a tabletop device, has been very explored for elder users in several domains. This interaction technique is an alternative to reducing the obstacles that older adults face in the use of computer systems, e.g., handling of peripherals. Many design guidelines are proposed in the literature for a wide range of products and systems for elders, e.g. websites, TV user interfaces. However, there is a lack of set of design guidelines and design recommendations of multi-touch interfaces that matches elder's needs. This paper presents a set of design guidelines and design recommendations distilled and extracted from most relevant works on design of multi-touch interfaces for elders available in the literature. The results are a set of design guidelines, useful for designers, application developers, usability specialists and researchers.**

Keywords-Human-Computer Interaction; Natural User Interfaces; Multi-Touch Interfaces; Design Guidelines; Elderly

I. INTRODUCTION

Globally, we can verify that the population is aging [1]. Mamolo and Scherbov [2] revealed that overall population is decreasing with a substantial increase of the elderly population. The number of people aged over 65 years in the world is predicted to be about 1 billion by 2030, a significant growth in comparison with 420 million verified in 2000 [3], registering the largest number ever. Zaphiris et al. [4] referred that in 2030, one habitant out of four will have above the age of 65.

The growing number of elders makes evident the need of development of technologies for this user group [5][6] that accommodate their needs [4][7].

This user group presents own characteristics that differ from other user groups [8], part of them are age-related changes, e.g., physical and cognitive changes [9]. Nevertheless, there are in the literature several studies researching on age-related changes and how they influence the design of user interfaces [1][5][6][10][11] and others that focused the relationship of older adults with technologies [26][29].

The interaction through of traditional input devices, such as mouse and keyboard, is difficult for elders [7][12], and other interaction techniques have been explored to surpass these limitations [9], e.g., using multi-touch interfaces [7] [9][13] as a more natural user interface [14].

The design of a touch-based user interface should suit elder's needs [11] to be easily used [5], becoming an elder-friendly user interface.

A considerable amount of works has been published on design guidelines for older adults, e.g., for design of websites [4][15], design of mobile user interfaces [16][17], design of TV user interfaces [8][18]. These studies usually propose a list of guidelines for designing and evaluating the user interfaces. This is not verified in works with a list of design guidelines of multi-touch interfaces, which are scarce in the literature.

Although the design guidelines of other types of user interfaces can be used partly in the design of a multi-touch interface, some characteristic aspects are not addressed, e.g., the identification of the most suitable gestures for this target audience.

In that sense, this paper contains a set of design guidelines and design recommendations refined and extracted from most relevant works on design of multi-touch interfaces for elders existing in the literature.

The rest of the paper is structured as follows. Section II reviews the background of this work, including, age-related changes, multi-touch interfaces, as well as designing multi-touch interfaces for elders. Section III describes the methodology used. Section IV provides a review of design guidelines of multi-touch interfaces for elders. Section V presents the final set of design guidelines. Finally, Section VI concludes and presents future research directions of this paper.

II. BACKGROUND

In this section, we briefly outline the main concepts associated with this work, including, age-related changes, followed by an overview of suitable input devices and multi-touch interfaces for elders and concludes with considerations regarding the design of multi-touch interfaces for elders.

A. Age - Related Changes

The aging process is typically accompanied with physical changes (namely visual, auditory and motor changes) and cognitive changes (decline in memory and attention) [10][19][20][21].

Visual problems are noticed around the age of 40 [6]. The visual capabilities of a person are affected with aging, being the most significant changes observed in the visual

 ISBN: 978-1-61208-325-4

acuity [22], presbyopia [21], peripheral vision [21] and dark adaptation [21][22].

Hearing loss is verified in approximately 50% of all men over the age of 65 and 30% of women [6]. The human ear can hear sounds from the range of frequencies from 20Hz to 20000Hz. Aging causes a decrease in hearing [21], resulting in a loss of the capability to detect tones in all the frequencies. However, this loss is more recognized in the high-pitched sounds [23], where some elders do not recognize sounds with frequencies higher than 2500Hz [24].

With respect to motor changes, these are caused by a loss of muscle mass and flexibility [6]. The main changes are specifically, gait disturbances (i.e., immobility), balance difficulties (i.e., instability), and certain motor control problems (i.e., tremor) [25] and arthritis [12].

Memory is a multi-component system that combines aspects of storage and processing [26]. Normal aging, produces different degrees of decline in the several forms of memory [20], namely short-term memory (i.e., working memory) used for example in learning and interacting with new devices and long-term memory (i.e., permanent memory) used to store information over a long period of time.

Finally, attention that consists in the ability to focus on the items needed to perform a certain task [20][22] is also affected by aging. Individuals over 60 years old, have a substantial difficulty in processing complex tasks [27], and have problems maintaining attention span for long periods of time [28].

The use of touchscreens through multi-touch interfaces can accommodate largely these age-related limitations [6]. We will briefly summarize the most appropriate input devices for elders.

B. *Suitable Input Devices for Elders*

Human-computer interaction enables the use of a wide variety of input devices, such as keyboards, mice, touchpads, touchscreens [10]. These input devices can be divided in two categories, direct and indirect input devices [29]. A direct input device is characterized by direct user input on a display [10], e.g., touchscreen. An indirect input device needs coordinate spatial information, hand-eye coordination, and finger dexterity to operate the device and interacting with the user interface [30], e.g., touchpad.

Touchscreen devices reduced cognitive and coordination demands [29], being referred by research community as appropriated and preferred for elder users [6]. However, these devices present some disadvantages, such as, users hands may obscure the screen, the risk of inadvertent activation, as well as difficulties in the precision tasks [31].

Several interaction techniques are possible in the touchscreens devices [30], e.g., single-touch or multi-touch interaction [32].

In the single-touch interaction, only a point of contact is recognized, e.g., using a finger. This way of interaction allows performing basic operations, e.g., open and close programs or pushing buttons. On the other hand, most recent devices enable multi-touch interaction, detecting multiple simultaneous touch points [33], e.g., using the fingers of a single hand or the both hands for interacting with the user interface, through of certain gestures on surface.

Comparisons of touchscreens with other input devices are often referred by the scientific community [29][34][35].

Wood et al. [29] analyzed the performance of older adults in simple drag-and-drop tasks, using four different input devices, namely, touchscreen, enlarged mouse (EZ Ball), mouse and touchpad. As performance measures was used accuracy and time to complete parts of a game. Additionally, it was held a set of measurements, to assess visual memory, visual perception, motor coordination, and motor dexterity of the users. The results showed that mouse was the device that demanded greater effort to accomplish the proposed tasks.

Findlater et al. [34] studied the psychomotor performance between young adults and older adults using desktops and touchscreens. The evaluated tasks were pointing, dragging, crossing and steering in both devices and pinch-to-zoom in the touchscreens. The results showed that older adults have slower performance than younger adults, however, the use of touchscreen reduced the time in comparison with desktop and mouse in the tasks performed by older adults.

Schneider et al. [35] presented a study to compare the performance of different input devices, namely, mouse, touchscreen, eye-gaze input, and a hybrid interface, composed by eye-gaze input with other input devices. The better results were achieved using touchscreen and worst results using mouse, in the group of the elder users.

In addition to comparisons with other input devices, the touchscreens allow innumerous interaction techniques that are also focused in the literature, e.g., Motti et al. [30] presented a review on interaction methods using touchscreens by older adults.

The multi-touch interaction capabilities of touchscreens enable the use of sophisticated multi-touch interfaces that will be addressed in next section.

C. *Multi-Touch Interfaces for Elders*

Natural User Interfaces (NUI) were formed to establish new natural ways of communication between users and computer systems [14]. The term NUI enables the manipulating of a user interface in natural and intuitive form for human being [9][36]. NUI can be developed using the natural capabilities of humans, such as, touch, gestures, speech, facial expressions, body language, eye-gazing, or combining several input modalities designated by multimodal interfaces.

Multi-touch interfaces are a type of NUI, in which multiple simultaneous touch points on a user interface are recognized [33], allowing direct mapping of the input in the user interface [13][37], considered 'natural' and 'intuitive' their use [6][7][14].

The use of multi-touch interfaces has been receiving an increasing attention in recent time, given the diversity of devices that supports multi-touch and gestural input [38], such as mobile devices [31], tablet devices [13], tabletop devices [39], becoming these type of interfaces very popular [7], promoting easier hand-eye coordination [6], and interesting to users of all ages [6].

Multi-touch interfaces using the fingers of one or two hands were evaluated through different gestures on touchscreen, e.g., tap, drag, rotate, resize [11][32][34][40].

Piper et al. [40] showed that older adults are appropriate to performing multi-touch gestures, however, with difficulty in gestures that involve fine motor movements, i.e., gestures involving two or more fingers, e.g., rotate gesture. Leonardi et al. [11] mentioned that tap gestures using a single finger, are easy to understand and remember by older adults but with difficulties in drag gesture, due to lack of constant pressure. Findlater et al. [34] verified the presence of some errors in the "zoom out" gesture and the absence of errors in the "zoom in" gesture. Stößel et al. [31], referred that older adults are slower than young adults in multi-touch interaction, but with similar error-prone.

Tabletop devices are a type of touchscreens, very appropriate for elder's, making use of a large interaction area and with multi-user support [37][39], a playful way for promoting their face-to-face social interaction, like their daily activities, e.g., playing board games.

Gaming is an interesting domain for multi-touch interfaces. An illustrative example is [37], that explored multi-touch interfaces as a gaming platform for older adults.

The designing of multi-touch interfaces for elders requires the consideration of a diversity of aspects that will be addressed in the next section.

D. Designing of Multi-Touch Interfaces for Elders

As mentioned before, tabletop devices are appropriate for elder users, and the designing of multi-touch interfaces for this type of devices will now be described.

It is well known that aging inevitably brings changes to the physical and cognitive abilities of humans [41] and consequently the design of a user interface should accommodate the limitations caused by these changes [7], stimulating the development of technologies more usable by older adults [6].

Lists of design guidelines of multi-touch interfaces for elders are not abundant in the literature, however, Boustani [7] presented a set of touch-based design guidelines for elders, that is a good starting point for designing of multi-touch interfaces for elders, however some important aspects are missing, e.g., the reference of gestures desirable and avoidable by this type of users and guidelines to help the interface testing.

Although, in multi-touch interface design be possible the use of guidelines of other types of user interfaces, e.g., web design guidelines, given that, some guidelines are similar in both user interfaces, such as content layout design, text design, use of colors. However, specific aspects of multi-touch interfaces, e.g., the adequacy of gestures, ideal display size and feedback are not covered and are crucial in the design of an elder-friendly multi-touch interface. Designing of multi-touch interfaces for elders should aim at attenuating the limitations caused by aging. To this end, some practical guidelines will be mentioned according to age-related change:

- Visual changes - is suggested that the user interface should have an appropriate size of design elements and text [6][13], making use of high contrast colors [1] [6][7][8][18];
- Hearing changes - is recommended the presence of a control to adjust audio [1][6][8][20][21];
- Motor changes - the design of user interface should contains large targets for accurate selections [6][7] [8], avoid the use of scrolling [6][7][8][18] and exhibit slower response times [6][40];
- Memory changes - the interface should have appropriate feedback [1][42], including the current location in the system [7][8][18] and the use of meaningful icons [1][7][8][18];
- Attention changes - potential distracting elements should be avoided, such as, animations [7][8] and irrelevant informations [6][7][8][18].

In addition to the last practical guidelines mentioned, other considerations should be taken into account, e.g., tap gestures are easy to understand and remember being appreciated by elderly people, as well as iconic gestures [11]; positioning at the surface is critical, the user should be able to reach the entire interface and tactile user feedback and natural affordances are also needed [42].

The reviewing of works on design guidelines of multi-touch interfaces for elders is more detailed in Section IV.

III. METHODOLOGY

The list of design guidelines was reached using the methodology described in the following steps.

1) Selection of relevant works on design guidelines of multi-touch interfaces for elders: From literature were selected a set of most relevant works that focused on design guidelines of multi-touch interfaces for older adults.

2) Creation of an initial set of design guidelines: From the analysis of the works identified in step 1, it was created an initial set of 138 design guidelines.

3) Review, grouping and organization of the initial set of design guidelines: The initial set of design guidelines was reviewed, in order to: discover and associate identical guidelines, detect and resolve the guidelines that are in divergence and rewrite indistinct guidelines. After reviewing guidelines, 10 meaningful groups of guidelines were created, associating each guideline with a group, originating the final list of design guidelines.

4) Completing data of each design guideline: For each design guideline were filled the following fields: guideline number, guideline title, guideline group, guideline description, an illustrative example, the works that referred the guideline denominated guideline source and a set of tags that classifies the guideline.

5) Final set of design guidelines: The final set of design guidelines consists of a list of 113 grouped design guidelines, focusing the essential aspects on design of multi-touch interfaces for elders. The goal is to be a useful resource for designers, application developers, usability specialists and researchers.

IV. Review of Design Guidelines of Multi-Touch Interfaces for Elders

In this section, a selection of the most relevant works on design guidelines of multi-touch interfaces for elders available in the literature, will be presented and summarized.

Leonardi et al. [11] conducted a preliminary study on a touch-based gestural interface (Mobitable), to assist elders in the use of social networking. The study verified the appropriateness of this type of interfaces, mentioning some considerations on design, e.g., tilted or adjustable display, can ease the interface visualization.

Apted et al. [39] described the design of SharePic - a collaborative digital photograph sharing application. An easy to learn and remember multi-touch and gestural application used by multiple users on a tabletop. An evaluation was conducted, with young and older adults, verifying difficulty in two-handed gestures, performed by older adults. Design guidelines were pointed, e.g., should be possible to enlarge the interface elements and the interactivity should be focused on learnability and memorability.

Kin et al. [43] studied the performance in multiple target selection using a mouse, a multi-touch workstation, with a single finger, two fingers (one of each hand) and multiple fingers interaction. The results revealed greater speed with the use of multi-touch interaction, independently of the number of targets. Moreover, it was mentioned some design considerations of applications whose target selection is the primary task, e.g., the use of two-fingers in the multi-touch interaction does not provide any performance improvement in the multi-target selection.

Piper et al. [40] examined accessibility issues in the use of surface computing by older adults and investigated the appeal of this interaction technique for health care support. A study that involved the performing of set of tasks by older adults was done. The results showed that participants had difficulty with gestures involving two or more fingers, e.g., resize and rotate. Additionally, it was referred some design guidelines, such as, provide cues for interaction and the display size may be intimidating.

Silva and Nunes [5] conducted usability tests with older adults in the context of the European project Enhanced Complete Ambient Assisted Living Experiment (eCAALYX). From their experience a set of guidelines of usability tests for older adults was presented. These include guidelines, such as informing the older adult of the goal of the project beforehand, talk to privileged informers, and make it clear that they are not being tested.

Bachl et al. [42] discussed eight challenges in designing multi-touch interfaces, grouped in three categories, specifically, screen-based, user-based and input-based challenges. Some examples of challenges debated were: the necessity of tactile user feedback during the interaction, individual differences of the users, for instance hand size are relevant in designing of multi-touch interfaces and distinguishing and identifying users in the multi-user support.

Nunes [8] presented a set of design guidelines to guide the design and planning of usability tests of TV-based user interfaces for older adults, derived from the knowledge of the analysis, design and evaluation of a TV system for the European project eCAALYX [18]. Guidelines, such as, give them time to learn, use simple phrasing, use a very large font type, remove sound distractions and provide a good navigation are examples included in this collection.

Caseiro [13] created an Android tablet gaming platform to provide cognitive games. User management and the cognitive performance of players were the main features of this platform. Design considerations highlighted, use of big button size; use of colors with good contrast; use a suitable text size, among others.

Banovic et al. [44] explored the efficient design of context menus manipulated by a single hand on multi-touch surface. A context menu design for a horizontal tabletop surfaces was proposed, as well as, design guidelines for single hand multi-target selection using multi-touch interactions, e.g., interfaces should encourage users to approach from south (S), southeast (SE), or east (E) to the primary target and the use of index finger (index-anchored) or the thumb (thumb-anchored) to select the primary target does not influence the performance.

Boustani [7] presented a list of twenty nine guidelines distributed by ten groups of touch-based interfaces design guidelines for the elderly people, derived from a set of related works available in the literature. Examples of guidelines referred, irrelevant information on the screen should be avoided and blue and yellow or red and green tones should not be used. The identified guidelines were used in the design of application Keep in Touch (Kit) a platform easy to learn and understand in the communication between older adults.

Caprani et al. [6] investigated the use of touchscreen technology for elders. The research was focused on several aspects, such as, study of main characteristics of the older user, use of technology by older adults, comparison of touchscreens with other pointing devices, design guidelines for touch-screen devices, among others. Examples of referenced guidelines in this work: use high contrast between the elements of the user interface, use of buttons with big size and scrolling should be avoided.

Farage et al. [1] summarized the main age-related changes, addressing specifically, changes in the visual function; hearing; touch and temperature perception, mobility, and balance; memory and cognition. For each one age-related change were suggested concrete design guidelines to attenuate these limitations, e.g., use colors with good contrast (visual changes), increase duration of sound signal (hearing changes) and use of icons along with labels (cognitive changes).

Jin et al. [45] investigated the spacing and size of buttons in a touchscreen user interface used by older adults. During the experiments it was measured the reaction time, accuracy and user preferences. The results of this research showed that large button have shorter reaction times and larger space between buttons does not improve performance.

Nunes et al. [18] presented a set of thirteen recommendations for designing TV user interfaces for older adults. These recommendations were based on design, testing, and development of a TV-based health system. Use consistency to facilitate recognition, show the current selection clearly and give users time to read, are examples of recommendations suggested.

All works detailed here were used to form the list of design guidelines, presented in the next section.

V. FINAL DESIGN GUIDELINES

The final design guidelines are composed by 113 distinct guidelines, grouped under 10 distinct category headings, as illustrated in Figure 1. The categories used in the grouping of guidelines was based on classification used by Boustani [7] and Kurniawan and Zaphiris [46]. Some selected examples of each category are listed below.

Figure 1. Guidelines Categories [47]

G1 - Target Design

- Ensure the user can easily make interface elements larger (adjustable);
- Different physical properties have to be considered while designing the interface (e.g. size of buttons);
- Provide a cursor showing clearly the selected target. It should be obvious to older adults what can be selected and what cannot.

G2 - Use of Graphics

- Use icons along with labels. Icons should be simple and meaningful; text incorporated with the icon when possible;
- Use high contrast between the elements of the user interface. A high contrast between the foreground and background should exist;
- Blue and yellow or red and green tones should be avoided. Warm colors are the most suitable.

G3 - Navigation and Errors

- Provide a good navigation;
- Show the current location clearly;
- Design error messages that make it clear that the user is not the cause of the error;
- Make it easy for user to correct input errors.

G4 - Content Layout Design

- Concentrate information on the center of the screen;
- Avoid the use of scroll;
- Maintain consistency in the user interface. Screen layout, navigation and terminology used should be simple, clear and consistent;
- Remove user interface elements calling attention as soon as they are not needed.

G5 - User Cognitive Design

- Be prepared for older adults that refuse to learn;
- Make use of behaviors developed by older adults to cope with memory loss;
- Give them time to read. Older adults usually read more slowly (than younger adults);
- Avoid forcing users to read at very close distances.

G6 - Audio

- Enable older adults to adjust the volume at their will;
- Increase duration of sound signals;
- Use male voices for delivering auditory information;
- Remove sound distractions.

G7 - Text Design

- Use a very large font type;
- Use left-aligned text;
- Use an easy to read font family;
- Use medium or bold face type, e.g., Sans Serif type font, i.e., Helvetica, Arial. Avoid other fancy font types.

G8 - User Feedback and Support

- Accessibility issues should be taken into account, for example giving tools to allow the use by blind people;
- The lack of tactile user feedback also affects the user experience of data input on multi-touch interfaces;
- Use supporting peripherals if needed.

G9 - Multi-Touch Interaction

- Tap gestures (when applied to well recognized objects) are the easiest ones to understand and remember;
- Iconic gestures are very engaging;
- Provide cues for interaction for initial learning and sub- sequent use of the technology;

 ISBN: 978-1-61208-325-4

- Positioning at the surface is critical; the user should be able easily reach the interface corners.

G10- Interface Testing

- Inform the older adult of the goal of the project beforehand;
- Keep the test short and make use of breaks;
- Respect the opinions of the test participants.

Due to space restrictions for this paper, certain guidelines were omitted, as well as, descriptive data, including, guideline description, works that cited the guideline, illustrative examples, etc. The complete list of design guidelines can be obtained from [47].

VI. Conclusion and Future Work

In this paper, we presented a properly organized set of design guidelines of multi-touch interfaces for elders, refined and extracted from most relevant works presented in the literature. This set of design guidelines was structured in a very detailed and comprehensive way, covering the main age-related changes that might affect the usability of the interactivity in the multi-touch interfaces, becoming an important resource for designers, application developers, usability specialists and researchers to guiding and evaluating the design of this type of user interfaces for elder users.

As future work, we are considering the possibility of incorporating the design guidelines in an automatic detection system that identifies and suggests guidelines during the testing of a multi-touch interface; increasing the list of design guidelines with inclusion of other design guidelines; the reviewing and rating of the design guidelines by experts, in order to validate their usefulness, as well as a comparison of the design of a multi-touch interface for elders with and without the use of proposed guidelines to verify its impact on usability.

Acknowledgment

This work is being developed within the context of DEI/FEUP's doctoral programmme, ProDEI.

References

[1] M. A. Farage, K. W. Miller, F. Ajayi, and D. Hutchins, "Design principles to accommodate older adults," Glob. J. Health Sci., vol. 4, no. 2, Feb. 2012, pp. 2–25, doi: 10.5539/gjhs.v4n2p2.

[2] M. Mamolo and S. Scherbov, "Population Projections for Forty-Four European Countries : The Ongoing Population Ageing," Eur. Demogr. Res. Pap., 2009, pp. 63.

[3] O. Ribeiro, L. Fernandes, H. Firmino, M. R. Simões, and C. Paúl, "Geropsychology and psychogeriatrics in Portugal: research, education and clinical training.," Int. psychogeriatrics IPA, vol. 22, no. 6, Sep. 2010, pp. 854–863, doi:10.1017/S1041610210000347.

[4] P. Zaphiris, M. Ghiawadwala, and S. Mughal, "Age-centered research-based web design guidelines," in Proceedings of CHI EA '05 CHI 05 extended abstracts on Human factors in computing systems CHI '05, Apr. 2005, pp. 1897–1900, doi: 10.1145/1056808.1057050.

[5] P. A. Silva and F. Nunes, "3 x 7 Usability Testing Guidelines for Older Adults," in Proceedings of the 3rd Mexican Workshop on Human Computer Interaction, MexIHC'2010, Nov. 2010.

[6] N. Caprani, N. E. O'Connor, and C. Gurrin, "Touch Screens for the Older User," in in Assistive Technologies, Intech., F. A. Cheein, Ed. Ireland, Mar. 2012, pp. 104–128, doi: 10.5772/38302.

[7] S. Boustani, "Designing touch-based interfaces for the elderly," University of Sydney, Oct. 2010, pp.93.

[8] F. Nunes, "Healthcare TV Based User Interfaces for Older Adults," University of Porto, Jun. 2010, pp.181.

[9] B. Loureiro and R. Rodrigues, "Multi-touch as a Natural User Interface for elders: A survey," in Proc. of 6th Iberian Conference on Information Systems and Technologies CISTI 2011, Jun. 2011, pp. 573 –578.

[10] D. Taveira and S. D. Choi, "Review Study of Computer Input Devices and Older Users," Int. J. Hum. Comput. Interact., vol. 25, no. 5, Jul. 2009, pp. 455–474, doi:10.1080/10447310902 865040.

[11] C. Leonardi, A. Albertini, F. Pianesi, and M. Zancanaro, "An exploratory study of a touch-based gestural interface for elderly," in Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, Oct. 2010, pp. 845–850, doi:10.1145/1868914.1869045.

[12] S. J. Czaja, "Computer Technology and the Older Adult," in in Handbook of Human-Computer Interaction, M. G. Helander, Ed. Amsterdam: North-Holland, Aug. 1997, pp. 797–812.

[13] J. Caseiro, "Using the Android Tablet to Develop a Game Platform for Older Adults," University of Porto, Jul. 2011, pp. 127.

[14] D. Wigdor and D. Wixon, Brave NUI World: Designing Natural User Interfaces for Touch and Gesture. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc, Apr. 2011, p. 253.

[15] K. Pernice and J. Nielsen, "Web Usability for Senior Citizens: Design Guidelines Based on Usability Studies with People Age 65 and Older," Nielsen Norman Group, 2005,pp.126.

[16] M. S. Al-Razgan, H. S. Al-Khalifa, M. D. Al-Shahrani, and H. AlAjmi, "Touch-Based Mobile Phone Interface Guidelines and Design Recommendations for Elderly People: A Survey of the Literature," in Proceedings of 19th International Conference, ICONIP 2012, Part IV, 2012, Nov. 2012, pp. 568–574, doi:10.1007/978-3-642-34478-7_69

[17] M. Pattison and A. Stedmon, "Inclusive design and human factors : designing mobile phones for older users," Hum. Factors, vol. 4, no. 3, Dec. 2006, pp. 267–284.

[18] F. Nunes, M. Kerwin, and P. A. Silva, "Design recommendations for tv user interfaces for older adults: findings from the eCAALYX project," in Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility, ASSETS'12, 2012, Oct. 2012, pp. 41–48, doi:10.1145/2384916.2384924.

[19] N. Wagner, K. Hassanein, and M. Head, "Computer use by older adults: A multi-disciplinary review," Comput. Human Behav., vol. 26, no. 5, Sept. 2010, pp. 870–882, doi: 10.1016/j.chb.2010.03.029.

[20] D. Hawthorn, "Possible implications of aging for interface designers," Interact. Comput., vol. 12, no. 5, Apr. 2000, pp. 507–528, doi:10.1016/S0953-5438(99)00021-1.

[21] A. D. Fisk, W. A. Rogers, N. Charness, S. J. Czaja, and J. Sharit, Designing for older adults: Principles and creative human factors approaches, vol. 3, no. 3. CRC Press, 2009, p. 232.

[22] N. Charness and K. S. Warner, Impact of Technology on Successful Aging, 1st edition. Springer Publishing Company, Sep. 2003, p. 352.

[23] F. Schieber, "Aging and the senses," in in Handbook of Mental Health and Aging,, J. E. Birren, R. B. Sloane, and G. D. Cohen, Eds. San Diego, CA: Academic Press, 1992.

[24] L. N. D. Huey R.W., Buckley D.S., "Audible performance of smoke alarm sounds," in Proceedings of the Human Factors and Ergonomics Society, 38th Annual Meeting,, Jul.1994, pp. 147–151, doi:10.1016/0169-8141(95)00030-5.

[25] L. J. Schut, "Motor system changes in the aging brain: what is normal and what is not.," Geriatrics, vol. 53, no. 1, Sep. 1998, pp. 16–19.

[26] J. McCabe and M. Hartman, "Examining the locus of age effects on complex span tasks.," Psychol. Aging, vol. 18, no. 3, Sep. 2003, pp. 562–572.

[27] E. Commodari and M. Guarnera, "Attention and aging.," Aging Clin. Exp. Res., vol. 20, no. 6, Dec. 2008, pp. 578–584.

[28] M. Vercruyssen, "Movement control and the speed of behavior," in Hand-book of Human Factors and the Older Adult, W. A. R. (Eds. . A.D. Fisk, Ed. San Diego: Academic Press, Oct. 1996, pp. 55-86.

[29] E. Wood, T. Willoughby, A. Rushing, L. Bechtel, and J. Gilbert, "Use of computer input devices by older adults," J. Appl. Gerontol., vol. 24, no. 5, Nov. 2005, pp. 419–438, doi: 10.1177/0733464805278378

[30] L. G. Motti, N. Vigouroux, and P. Gorce, "Interaction techniques for older adults using touchscreen devices: a literature review," in Proceedings of the 25ième conférence francophone on l'Interaction Homme-Machine, Nov. 2013, pp. 125–135, doi:10.1145/2534903.2534920.

[31] L. T. M. Stößel, C., Wandke, H., and Blessing, "Gestural Interfaces for Elderly Users: Help or Hindrance?," Gesture Embodied Commun. Hum. Comput. Interact., vol. 5934, Feb. 2010, pp. 269–280, doi:10.1007/978-3-642-12553-9_24.

[32] G. Lepicard and N. Vigouroux, "Comparison between single-touch and multi-touch interaction for older people," Lect. Notes Comput. Sci. - Comput. Help. People with Spec. Needs, vol. 7382, no. PART 1, Jul. 2012, pp. 658–665, doi: 10.1007/978-3-642-31522-0_99.

[33] A. Teiche, A. K. Rai, C. Yanc, C. Moore, D. Solms, G. Çetin, J. Riggio, N. Ramseyer, P. D'Intino, L. Muller, R. Khoshabeh, R. Bedi, M. T. Bintahir, T. Hansen, T. Roth, and S. Sandler, Multi-Touch Technologies. NUI Group, 2009, p. 100.

[34] L. Findlater, J. E. Froehlich, K. Fattal, J. O. Wobbrock, and T. Dastyar, "Age-Related Differences in Performance with Touchscreens Compared to Traditional Mouse Input," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI' 2013, Apr. 2013, pp. 343–346, doi: 10.1145/2470654.2470703.

[35] N. Schneider, J. Wilkes, M. Grandt, and C. M. Schlick, "Investigation of Input Devices for the Age-differentiated Design of Human-Computer Interaction," in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 52, no. 2, Sep. 2008, pp. 144–148, doi:10.1177/1541931 20805200202.

[36] G. Glonek and M. Pietruszka, "Natural User Interfaces (NUI): Review," Jounal Appl. Comput. Sci., vol. 20, no. 2, 2012, pp. 27–45.

[37] T. Marques, F. Nunes, P. Silva, and R. Rodrigues, "Tangible interaction on tabletops for elderly people," in Proceedings of the 10th international conference on Entertainment Computing ICEC'11, Oct. 2011, pp. 440–443, doi: 10.1007/ 978 -3-642-24500-8_61.

[38] D. Wigdor, J. Fletcher, and G. Morrison, "Designing user interfaces for multi-touch and gesture devices," in Proceedings of the 27th international conference extended abstracts on Human factors in computing systems CHI EA 09, Apr. 2009, pp. 2755–2758, doi:10.1145/1520340.1520399.

[39] T. Apted, J. Kay, and A. Quigley, "Tabletop sharing of digital photographs for the elderly," in Proceedings of the SIGCHI conference on Human Factors in computing systems CHI '06, Apr. 2006, pp. 781–790, doi:10.1145/1124772.1124 887.

[40] A. M. Piper, R. Campbell, and J. D. Hollan, "Exploring the accessibility and appeal of surface computing for older adult health care support," in Proceedings of the 28th international conference on Human factors in computing systems CHI '10, Apr. 2010, pp. 907–916, doi:10.1145/1753326.1753461

[41] J. Häikiö, A. Wallin, M. Isomursu, H. Ailisto, T. Matinmikko, and T. Huomo, "Touch-based user interface for elderly users," in Proceedings of the 9th international conference on Human computer interaction with mobile devices and services MobileHCI 07, Sep. 2007, pp. 289–296, doi:10.1145/137 7999.1378021.

[42] S. Bachl, M. Tomitsch, C. Wimmer, and T. Grechenig, "Challenges for Designing the User Experience of Multi-touch Interfaces," in Proceedings of Engineering Patterns for Multi-Touch Interfaces workshop (MUTI'10) of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems, Jun. 2010, pp.365-370.

[43] K. Kin, M. Agrawala, and T. DeRose, "Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation," in Proceedings of Graphics Interface, GI '09, May. 2009, pp. 119–124.

[44] N. Banovic, F. C. Y. Li, D. Dearman, K. Yatani, and K. N. Truong, "Design of unimanual multi-finger pie menu interaction," in Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces ITS 11, Nov. 2011, pp. 120–129, doi:10.1145/2076354.2076378.

[45] Z. X. Jin, T. Plocher, and L. Kiff, "Touch screen user interfaces for older adults: button size and spacing," Lect. Notes Comput. Sci., vol. 4554, no. Universal Access in Human Computer Interaction. Coping with Diversity, Jul. 2007, pp. 933–941, doi:10.1007/978-3-540-73279-2_104

[46] S. Kurniawan and P. Zaphiris, "Research-derived web design guidelines for older people," in Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility Assets '05, Oct. 2005, pp. 129–135, doi:10.1145/1090785.1090810.

[47] B. Loureiro and R. Rodrigues, "Design Guidelines of Multi-touch Interfaces for Elders - Online Version," 2013. [Online]. Available: http://bit.ly/17ouqyF. [Accessed: 02-Feb-2014].

PolyPie: A Novel Interaction Techniques For Large Touch Surfaces With Extended Wall Displays

Ihab Maged *, Michael Louis *, Mohamed Thabet * and Ayman Atia †
* Faculty of computer science
MIU university , Cairo, Egypt
Email: ihab090178, michael.louis, mohamed.thabet @miuegypt.edu.eg
† HCI Lab, Faculty of computer science and information systems, Helwan University
Email: ayman@fci.helwan.edu.eg

***Abstract*—The paper presents new interaction techniques for large touch tables and large-wall projected screens. The extended projected screen is integrated as an extension to the touch table where the user can grab the menu items on the projected screen from the touch table or by interacting directly with the projected screen using hand gestures. The system proposes multi-touch controls and hand gestures that aim to minimize the user effort and body movements while interacting with the large touch table and the extended projected large-wall display. We present ”PolyPie” as a group of three touch interaction techniques for large display touch surfaces and two hand gestures techniques for interacting with the large-wall projected display. The proposed touch techniques are the Dynamic Pie Magnifier, Poly-Fingers Grab, Five Fingers Shadow Grab. We conducted a preliminary study for the proposed interaction techniques. The results showed that the proposed techniques helped the users to interact much easier with large display table and wall screen. The users were able to access distant files in acceptable time and more smoothly.**

Keywords-Large Display Interaction;Gesture Interaction;Multi-touch Tables;Interaction Techniques

I. INTRODUCTION

Large Display Screens are becoming widely used after recent progression of technology. Ball and North [1] showed that by launching several windows on a single large screen will facilitate different advantages for the users. The main benefit for working on those large displays is gaining the usability of previewing relatively large sets of data. The main challenge is to find out the proper interaction techniques that fit with interacting with those displays. Tan et al. [2] stated that large display screens save much time and effort when they are compared to the normal display devices especially when they are used in group-ware activities. Although those large displays may be perfectly used by a single user but by the help of certain interaction techniques. Also, the interaction techniques designed for traditional computers, such as the trackball mouse are not always appropriate for use with other form factors, especially for large displays. The traditional trackball mouse needs to be dragged from the bottom of the display up to the end in order to a grab or select a single menu item. Khan et al. [3] stated that in certain circumstances, the user spent a significant amount of time trying to find out where the mouse pointer was on the screen. The problem the users face is that it is impossible for them to reach distant windows and keep track of many launched windows. Vogel et al. [4] showed that in some certain scenarios, users have to step back or move from their place in order to access a specific item on a large display. There are some tasks that are needed to be handled from a distance in some specific environments, such as sorting slides, photos and pages spread over a large display surface or even a public place that contains a large mounted display that is needed to be accessed. Radloff et al. [5] showed that large displays in meeting rooms will assure a high level of presentations and interactions between the users.

In this paper, we proposed some new interaction techniques for working on a large touch table which is also extended to another projected large-wall display. The main concern is to embed the interaction techniques into the seamless interactions figured out by the user in order to gain the maximum usability and the minimum effort. Our motivation in this work is to prevent the user from being frustrated while working on these large extended displays, so the extended projected display and the touchable surface must be logically integrated and accessed as a single regular display with the proposed techniques. We conducted a preliminary experiment concerning the time of the proposed interaction techniques. We discussed the advantages of each technique and its usage in different situations and applications with large surfaces.

II. RELATED WORK

Parker et al. [6] showed that large display screens are now facing a problem in finding useful interaction techniques that will help its users to reach any far object. There are some techniques presented by researchers by using several input method like laser pointers. Myer et al. [7] showed that using laser pointers are problematic in terms of accuracy and speed. Moreover, Laser pointer are used more effectively on wall projected screens not a big display table.

Malik et al. [8] showed that there are several interaction techniques implemented for accessing far distant desktop items on a touchable surface to maximize the comfort of the user. They implemented their interaction techniques on a touch pad. A technique similar to the Vacuum [9] was implemented in which all the menu items between the angles of two fingers are dragged to each finger. Also, on touching the touchable surface with the five fingers, the nearest five menu items are dragged to each finger on the touch surface. Drag and drop is also implemented as if the user is catching a menu item and then leave the menu item to be dropped with his five fingers.

Bezerianos et al. [9] conducted the Vacuum technique, which works by drawing two projected lines till the end of the screen from a single source point, a polygon will be drawn and all the items covered in the polygon area will be dragged to the source point. The Vacuum technique was compared with the Drag-and-Pop technique and the experimental results showed that the vacuum technique is relatively faster and more accurate. The Vacuum technique has some limitations on the touch table especially on changing the angle between the two fingers. The user needs to remove his hand and hold it one more time on the touch table to change the angle.

Radloff et al. [5] discussed grouping of several large displays and how the user interacts with them. The user interacts with the projected displays using a Wii remote where he/she is able to manage and drag the menus and graphs instantly from one screen to another. The paper stated that the interactions with the projected screen by using a depth camera could be more appropriate than the Wii remote.

Kim et al. [10] showed that there are some interaction techniques used on the multi-touch screens where these techniques have to be predefined according to the number, position, and movement of the fingertips of the user. When the user touches the screen an image is captured and preprocessed to remove the noise and determine the movement of the fingers. The gesture commands were to manipulate the user interface. Bi et al. [11] conducted the MagicDesk that works on integrating touch capabilities with desktop work, which brings maximum accuracy and speed for the user as the keyboard and mouse technique is one of the fastest techniques but with integrating some touch interactions for desktop work introduces a new way for working on a desktop environment. The main contribution in the paper is using an interactive tabletop by interacting with touch techniques on the table and updating the monitor in real time.

Potter et al. [12] used Leap motion for Australian sign language. Leap Motion is a small usb device which can be placed on any physical surface , facing upward . It uses two monochromatic IR cameras and three infra-red LEDs , the device can detect motion to a distance of 1 meter (3 feet). The LEDs generate a 3D pattern of dots of IR light , the IR cameras then generates almost 300 frames per second of the reflected data. The leap motion controller software synthesize 3D position data by comparing the 2D frames which is generated by the IR Cameras. Marquardt et al. [13] showed that the main idea of unifying touch and gestures on and above a digital surface is to create a continuous interaction space where the user moves any tangible objects in the space and can also use touch and gestures techniques. The continuous interaction space can be illustrated as follows; a 3D area above the touch surface is conducted where the user interacts with his gestures without even touching the touch surface. For example, Mirrored Gestures for Redundancy is one of the techniques used where the person interacts with the touch surface using hand gestures instead of interacting directly on the surface. Kin touch is a way of integrating touch capabilities with Kinect [14] for facilitating the life of visually impaired people by using tactile or interactive maps. The software works on two applications, one for tracking fingers with Kinect image and the other for detecting the position of the fingers on the touch surface.

Shoemaker et al. [15] showed a new way for human computer interactions dealing with 5m X 3m large display wall depending on reality based interaction and whole body interfaces. The major contributions of their work was presenting a new body-centric approach specific to large display walls. The first step in their work was to sense the body by a vision tracker, then modelling the body position and at last analysing the interaction done by the user. Fikkert et al. [16] proposed new hand gestures for interacting with large-wall projected screens. They conducted their experiment on a large projected display where the task is to allocate a specific town on a large map. They introduced a set of gestures, such as the point, select, deselect and three other gestures for activating, deactivating and resizing.

Adrian et al. [17] implemented the SuperFlick, which is a technique for long-distance object placement on digital tables. SuperFlick is based on a remote Drag-and-Drop technique on a thrown object. The user does not need to wait until the flow is finished and the system knows the final position of the object as soon as it is thrown. Drag-and-Pop and Drag-and-Pick are techniques for accessing remote screen on touch and pen operated systems. Patrick et al. [18] showed that the users conducted some errors when the menu items are close to each other or when they are removed from their home screen. Doeweling et al. [19] implemented a new way for representing the Drag-and-Drop technique for large display touchable surfaces, which aims to optimize the short-backs of the regular drag and drop technique. A fully interactive proxy- targets operations are introduced and compared with the regular Drag-and-Drop technique.

III. System Overview

Our system is composed of a large touch surface and a large projected display integrated together, as shown in Figure 1. The user interacts with the projected screens using a Depth camera that is above the projected screen and in front of the user. The projected display is continuously extended and integrated to the touch surface. There are two scenarios for the user to grab the far away menu items. The first scenario, is working with the touch techniques to grab the far away menu items on the touch surface. However, this touch technique can work as an extension towards the projected display in order to grab all the menu items over the projected display and the touch surface. The second scenario is that the user grabs multiple items from the projected screen to the touch surface instantly using some hand gestures. A continuous interaction space is achieved between the projected display and the touch surface.

A. *Hardware Configurations*

The touch surface is created by Frustrated Total Internal Reflection (FTIR) technique used by the multi-touch community for building touch tables. A Stream of Infra-red LEDS is surrounded within an acrylic sheet where the Infra-red light enters the acrylic with a higher refractive index, at an angle of incidence greater than a specific angle. The FTIR method

 ISBN: 978-1-61208-325-4

Figure 1. Extended wall and large touch surface system overview

uses this principle of Total Internal Reflection in order to flood the Infra-red light inside the piece of acrylic. When the user touches the surface, the light rays are frustrated, since they can now pass through the contact material and the reflection is no longer total at that point. This frustrated light is scattered downwards towards an Infra-red camera in terms of blob positions where the camera software can track and analyze them.

A projector works as a displaying tool for the touch surface. Basically, the projector is inside the table and it is directed and positioned to a mirror. The mirror reflects the display with a higher index to the acrylic surface that is covered with a compliant layer. Our touch table dimensions are 150 X 120 X 100 *cm* where a single camera works on covering all the acrylic sheet. We used a single graphics card with multiple outputs for connecting the projector displaying the Projected-Wall screen into the same PC connecting the touch surface projector.

B. Software Configurations

Figure 2 shows the system's overview. It is composed of :; Input Layer, Intermediate Hardware Layer, Hardware Layer, Software Server Layer, Software Client Layer, Application Service Layer, Application Layer and Presentation Layer.

The Hand Gestures block is the Kinect Depth Camera input layer while the Touch Input block is the touch surface table input. The Intermediate Hardware Layer block contains the FTIR Touch table components that works for reflecting the IR light into the Camera. The Hardware Layer block includes the PlayStation 3 (PS3) Eye-Cam block that captures the reflected IR light and the Depth Camera block, which is the Kinect that capture the body skeleton of the user. We used the Community Core vision (CCV) as a Software Server Layer, which is the camera software that works on filtering the captured image and calibrating the touch input with the projected Screen. The TUIO protocol used for exchanging messages between the software client layer and the software server layer. The Microsoft Kinect Software Development Kit works for analyzing the users hand gestures. Gesture was used to select multiple items from projected display and other one for selecting all items on the projected screen.

The Application Service Layer is the Gestures Recognition Part, it implements the TUIO Client interface and the Microsoft

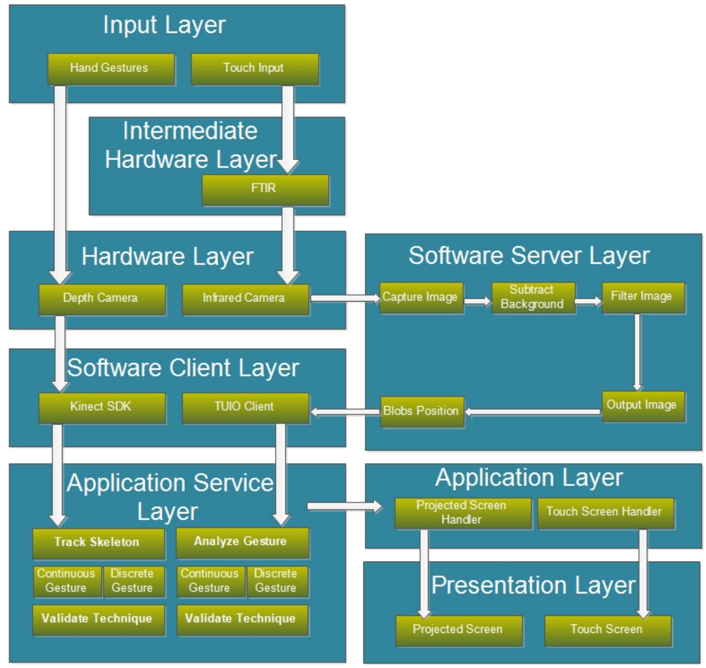


Figure 2. Block Diagram: System Overview

Kinect interface in order to start analyzing the gestures and compare the input gestures with predefined gestures. The Application layer works on animating and grabbing the selected items for the user's position based on the validated technique received by the Gesture Listener in the Application layer and displays the feedback to the presentation layer for the user.

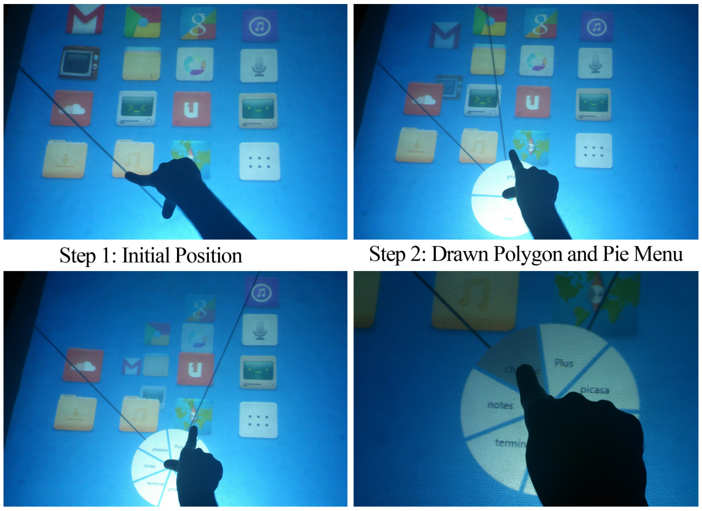
IV. Interaction Techniques

Interaction with large display screens requires some special interaction techniques in order to grab the far away menu items on the screen as it is impossible to select these items from the user's position. The user needs to move and stretch his arm towards the touch surface in order to select a specific item. The user cannot select the first top left item from his place. We implemented three touch interaction techniques for large display touch surfaces. The touch techniques are, the "Dynamic Pie Magnifier", the "Poly-Fingers Grab" and the "Five Fingers Shadow Grab". We used the Ray Tracing algorithm for all the techniques that works on checking the items inside a polygon.

A. Dynamic Pie Magnifier

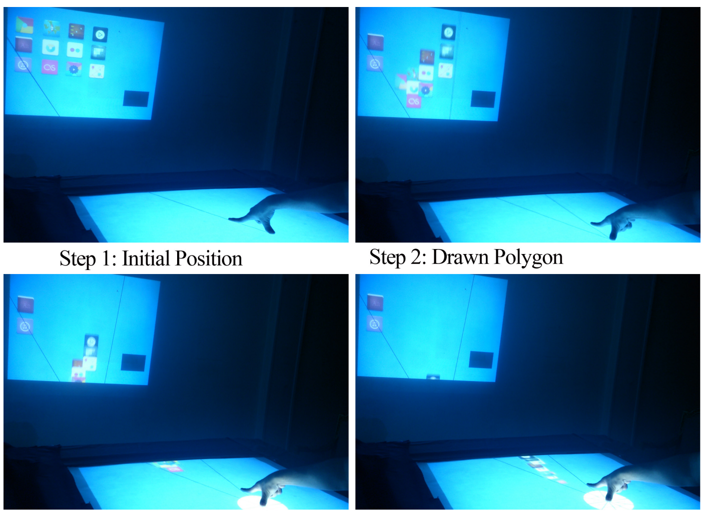
The Dynamic Pie Magnifier is basically based on interacting with the touch surface for grabbing multiple items from any position on the screen by using two fingers. The user holds down the thumb finger on the touch surface, then points with his second finger on the screen and start moving the second finger to the right position in order to maximize the size of the polygon. The shape of the drawn polygon is a cone polygon in order to insert the highest number of items possible.

The polygon area is maximized as long as the user drags his finger to the right and any allocated item inside the polygon will be moved and inserted into the Pie Menu. The Pie Menu appears on inserting the first item into the polygon. The



Step 1: Initial Position
Step 2: Drawn Polygon and Pie Menu
Step 3: Items Animated
Step 4: User Selects an Item

Figure 3. Touch surface: Dynamic Pie Magnifier Technique



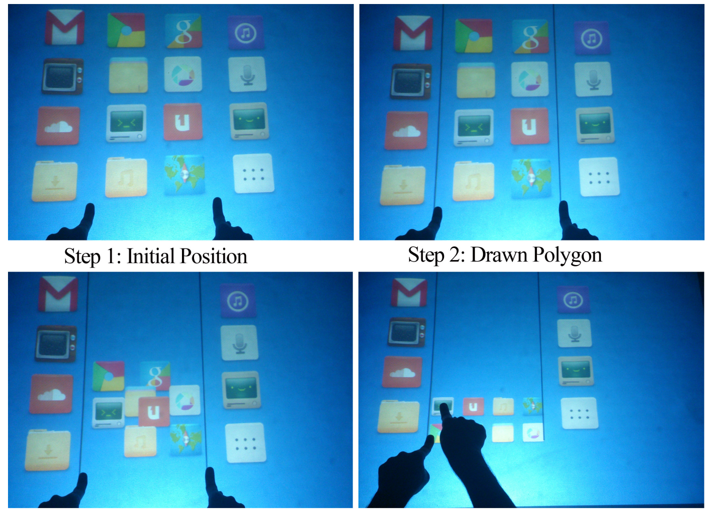
Step 1: Initial Position
Step 2: Drawn Polygon
Step 3: Items Animated to touch screen
Step 4: Items inserted into the Pie Menu

Figure 4. Extended Screen: Dynamic Pie Magnifier Technique

maximum amount of items that can be previewed on a Pie Menu are eight items. The user can swipe the pie menu in order to view more items. Figure 3 shows a single scenario for grabbing the far menu items to the user using the Dynamic Pie Magnifier Technique. The Dynamic Pie Magnifier technique can also work for grabbing the menu items on the Wall-Projected Screen to the touch surface as shown in Figure 4. The items are dragged from the projected display to the touch surface continuously and inserted into the Pie Menu.

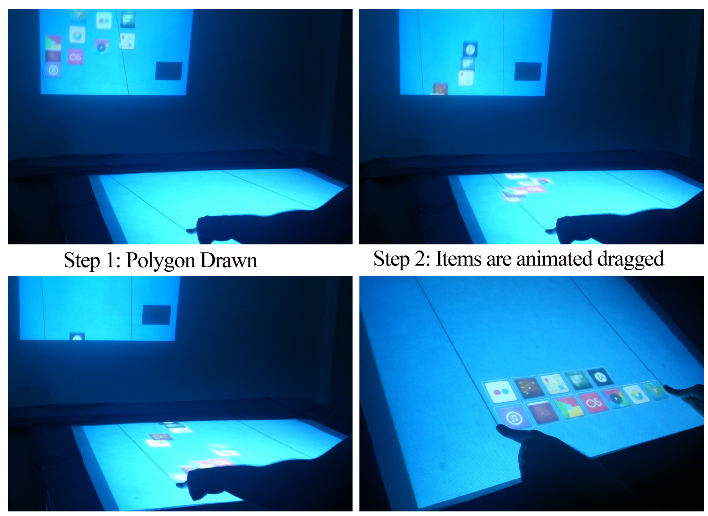
B. Poly-Fingers Grab

The Poly-Fingers Grab works on holding down the two fingers of two different hands on the touch surface. A straight line is drawn till the end of the touch surface from each finger in order to display a polygon for the user. The user can change the size of the polygon by dragging his two fingers to the left or right for maximizing and minimizing the polygon. The items inside the polygon are animated and dragged to the user's left finger position. Other menu items are grabbed to the user as long as the polygon is maximized.



Step 1: Initial Position
Step 2: Drawn Polygon
Step 3: Items Animated
Step 4: User Selects an Item

Figure 5. Touch surface: Poly-Fingers Grab Technique



Step 1: Polygon Drawn
Step 2: Items are animated dragged
Step 3: Continuous Animation
Step 4: Sorted Items at the user's position

Figure 6. Extended Screen: Poly-Fingers Grab Technique

The user can change the angle of the drawn lines by increasing the top position of one finger with a higher index than the other finger. On the other hand, all the items that are not anymore inside the polygon are dragged to their original position as long as the user minimizes the distance between the two fingers. Figure 5 shows a single scenario for grabbing the far menu items to the user using the Poly-Fingers Grab Technique. The Poly-Fingers Grab technique can also work for grabbing the menu items on the Wall-Projected Screen to the touch surface as shown in figure 6. The items are dragged from the projected display to the touch surface continuously where the items are sorted at the user's finger position. The user can finally select the target item with on of his fingers on the screen.

 ISBN: 978-1-61208-325-4

C. Five Fingers Shadow Grab

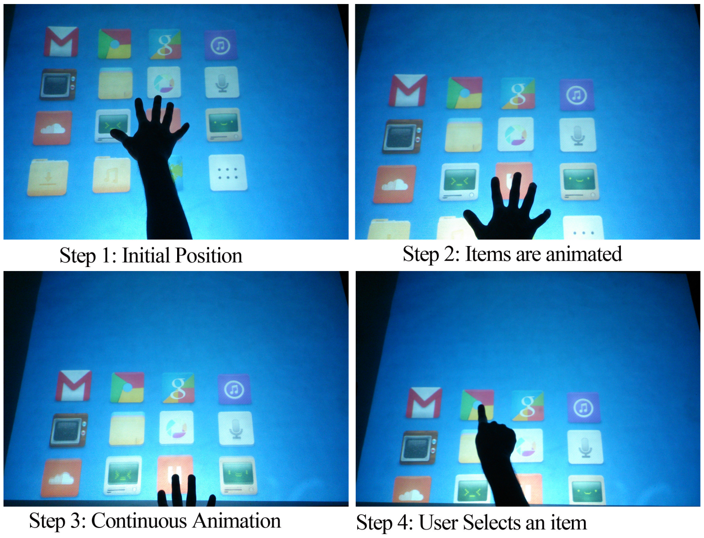


Figure 7. Touch surface: Five Fingers Shadow Grab Technique

The Five Fingers Shadow Grab technique simply works when the user puts his five fingers on the touch surface. The user starts grabbing his hand to the lower border of the table which is the user's position as shown in Figure 7. On the other hand, if the extended screen is activated, only the menu items on the projected screen will be dragged to the touch surface until they are close enough to the user. The technique works for grabbing all the items on the screen. The user finally selects one of the items from the grabbed items. The items are returned to their initial position if the user did not select any of the items for 2 seconds.

V. Experiments

Two Female and 8 Male volunteers participated on trying the interaction techniques over the touch surface where their age range is from 20 to 25 years old. They are all familiar with the traditional mouse input device but they have never interacted with a large touch screen or extended displays. We conducted a training session for each user before performing the real experiments. The experiment scenarios were fixed across all users to grab the most far menu items on the screen to the user's and select a target item (icons or files). We showed the participants 16 items and ask them to select 2 item from the first row. The users repeat each scenario 3 times.

A. Experiment 1 - Large Touch Surface

The goal of this experiment is to measure the time for the interaction techniques on grabbing the far away items on the touch surface. Furthermore, we compared the interaction techniques with the mouse device. The interaction techniques on the touch screen results are shown in figure 8. After conducting the experiments, it was shown that the Dynamic Pie Magnifier is the fastest technique and achieve a near result to the mouse device. The Five Fingers shadow Grab conducted more time as it was frustrating for the user to drag his five fingers together at the same time for long distance over the surface. The Mouse is faster than the other interaction techniques as the user does not need to grab the items to his position. Although, the mouse cannot be used as an input device for the touch surface, but we compared our techniques with the mouse input device to ensure that our techniques does not waste time while interacting with the touch surface. Moreover, the participants were more familiar with the mouse device.

B. Results 1

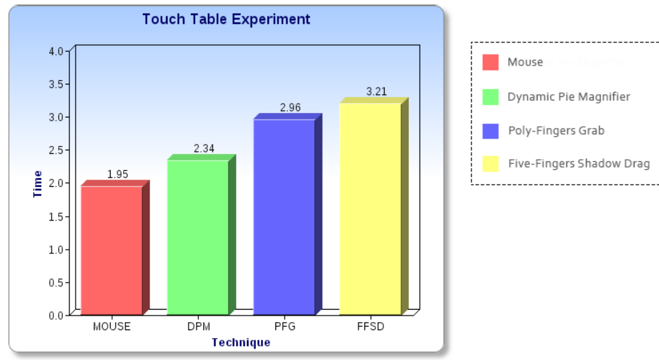


Figure 8. Large Touch Surface Interaction Techniques average time in seconds

The Dynamic Pie magnifier scored the best results after the mouse as all the selected items are inserted in a pie menu around the user's thumb. So, it was closer to the user and faster than the Poly-fingers grab, which works by grabbing the selected icons and listing them in rows and columns depending on the distance between the two fingers.

C. Experiment 2 - Extended wall display

The goal of this experiment is to test the interaction techniques on grabbing the projected display menu items to the touch surface. We compared all the techniques we developed in addition to mouse and hand gesture interaction. We tested two hand gesture techniques for the projected display. First, The "Multiple Selection" hand gesture that is based on the pause technique but the user has to raise his left hand in order to enter the selection mode. Second, the "Select All" Hand Gestures technique works for selecting all the items in the projected screen. The user has to raise his both hands in a crossing position, the user drags these items till the end of the screen coordinates and they are relatively dragged to the touch surface.

D. Results 2

The experiment shows that the user performance while interacting with the Poly-Fingers Grab is better than the Dynamic Pie Magnifier unlike the first experiment as shown in figure 9. The Poly-Fingers Grab was more accurate as the drawn straight polygon on the projected display was more imaginable to the user, while the cone polygon was not very familiar when it was extended to the projected display. The angle between the two drawing lines in the Dynamic Pie Magnifier relatively increases thus, lots of items are dragged to the touch surface which conducts higher time range on selecting the target item from the pie menu. On the other hand, the Poly-Fingers Grab drags the items directly to the touch surface and the user directly select the target.

 ISBN: 978-1-61208-325-4

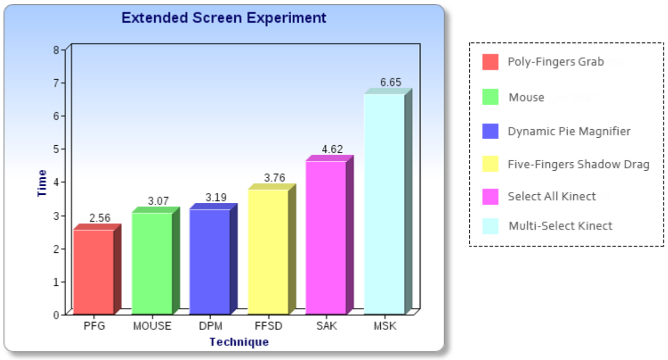


Figure 9. Extended Wall-Screen With Touch Surface Interaction Techniques Results average time in seconds

The Five Fingers Shadow Grab still conducts a higher time range for grabbing the items from the extended display and user movement over the surface was more longer. We have tested PolyPie with two different applications, first was live feedback presentation application for smart meeting rooms. We get some users feedback about feeling interested that they can access shared materials from the projected screen on their table interface smoothly. The second application was for sharing pictures between users sitting on the same large table. Users were found the pie menu interface easy to use and was fast way to access far items. Finally, the proposed interaction techniques will increase the productivity of large display touch tables.

VI. Conclusion and Future Work

We presented "PolyPie", an interaction techniques for large displays. Different touch interaction techniques are proposed in order to minimize the user effort while interacting with large display touch surfaces. The experiments showed that Dynamic Pie Magnifier achieved acceptable results while grabbing the items from the touch surface only. On the other hand, the Poly-Fingers Grab technique was better when the user grabs the items from the projected display to the touch surface. The Five Fingers Shadow Grab was frustrated while dragging the five fingers concurrently. Our future work is to create some techniques for grabbing the items from the touch table to the extended display.

References

[1] R. Ball and C. North, "Analysis of user behavior on high-resolution tiled displays," in Interact 2005. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 350–363.

[2] D. S. Tan, D. Gergle, P. Scupelli, and R. Pausch, "Physically large displays improve performance on spatial tasks," ACM Trans. Comput.-Hum. Interact., vol. 13, no. 1, Mar. 2006, pp. 71–99.

[3] A. Khan, J. Matejka, G. Fitzmaurice, and G. Kurtenbach, "Spotlight: directing users' attention on large displays," ser. CHI '05. New York, NY, USA: ACM, 2005, pp. 791–798.

[4] D. Vogel and R. Balakrishnan, "Distant freehand pointing and clicking on very large, high resolution displays," in Proceedings of the 18th annual ACM symposium on User interface software and technology, ser. UIST '05. New York, NY, USA: ACM, 2005, pp. 33–42.

[5] A. Radloff, A. Lehmann, O. Staadt, and H. Schumann, "Smart interaction management: An interaction approach for smart meeting rooms," in Proceedings of the 2012 Eighth International Conference on Intelligent Environments, ser. IE '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 228–235.

[6] J. K. Parker, R. L. Mandryk, and K. M. Inkpen, "Tractorbeam: seamless integration of local and remote pointing for tabletop displays," in Proceedings of Graphics Interface 2005, ser. GI '05. School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada: Canadian Human-Computer Communications Society, 2005, pp. 33–40.

[7] B. A. Myers, R. Bhatnagar, J. Nichols, C. H. Peck, D. Kong, R. Miller, and A. C. Long, "Interacting at a distance: measuring the performance of laser pointers and other devices," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '02. New York, NY, USA: ACM, 2002, pp. 33–40.

[8] S. Malik, A. Ranjan, and R. Balakrishnan, "Interacting with large displays from a distance with vision-tracked multi-finger gestural input," in UIST 2005, ser. UIST '05. New York, NY, USA: ACM, 2005, pp. 43–52.

[9] A. Bezerianos and R. Balakrishnan, "The vacuum: facilitating the manipulation of distant objects," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '05. New York, NY, USA: ACM, 2005, pp. 361–370.

[10] S.-G. Kim, J.-W. Kim, K.-T. Bae, and C.-W. Lee, "Multi-touch interaction for table-top display," in Proceedings of the 16th international conference on Advances in Artificial Reality and Tele-Existence, ser. ICAT'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 1273–1282.

[11] X. Bi, T. Grossman, J. Matejka, and G. Fitzmaurice, "Magic desk: bringing multi-touch surfaces into desktop work," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 2511–2520.

[12] L. E. Potter, J. Araullo, and L. Carter, "The leap motion controller: A view on sign language," in Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration, ser. OzCHI '13. New York, NY, USA: ACM, 2013, pp. 175–178. [Online]. Available: http://doi.acm.org/10.1145/2541016.2541072

[13] N. Marquardt, R. Jota, S. Greenberg, and J. A. Jorge, "The continuous interaction space: interaction techniques unifying touch and gesture on and above a digital surface," in IFIP TC 13 - Volume Part III, ser. INTERACT'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 461–476.

[14] A. Brock, S. Lebaz, B. Oriola, D. Picard, C. Jouffrais, and P. Truillet, "Kin'touch: understanding how visually impaired people explore tactile maps," in CHI '12 Extended Abstracts on Human Factors in Computing Systems, ser. CHI EA '12. New York, NY, USA: ACM, 2012, pp. 2471–2476.

[15] G. Shoemaker, T. Tsukitani, Y. Kitamura, and K. S. Booth, "Body-centric interaction techniques for very large wall displays," in Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, ser. NordiCHI '10. New York, NY, USA: ACM, 2010, pp. 463–472.

[16] W. Fikkert, P. van der Vet, G. van der Veer, and A. Nijholt, "Gestures for large display control," in Proceedings of the 8th international conference on Gesture in Embodied Communication and Human-Computer Interaction, ser. GW'09. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 245–256.

[17] A. Reetz, C. Gutwin, T. Stach, M. Nacenta, and S. Subramanian, "Superflick: a natural and efficient technique for long-distance object placement on digital tables," in Proceedings of Graphics Interface 2006, ser. GI '06. Toronto, Ont., Canada, Canada: Canadian Information Processing Society, 2006, pp. 163–170.

[18] P. Baudisch, E. Cutrell, D. Robbins, M. Czerwinski, P. Tandler, P. T, B. Bederson, and A. Zierlinger, "Drag-and-pop and drag-and-pick: techniques for accessing remote screen content on touch- and pen-operated systems." Microsoft Research, 2003, pp. 57–64.

[19] S. Doeweling and U. Glaubitt, "Drop-and-drag: easier drag & drop on large touchscreen displays," in Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, ser. NordiCHI '10. New York, NY, USA: ACM, 2010, pp. 158–167.

Trombosonic: Designing and Exploring a New Interface for Musical Expression in Music and Non-Music Domains

Oliver Hödl and Geraldine Fitzpatrick
Human Computer Interaction Group
Institute for Design and Assessment of Technology
Vienna University of Technology
oliver@igw.tuwien.ac.at, geraldine.fitzpatrick@tuwien.ac.at

***Abstract*—The "Trombosonic" is a new digital music instrument based on the foundational principles of the slide trombone. An ultrasonic sensor combined with a red laser allows the performer to play the instrument using similar movements to playing a trombone to change the pitch, by moving one hand back and forth even though there is no physical slider available. Furthermore, additional sensors enhance musical expression by gestural movement of the whole interface and by using the breath. Due to its compact size and the lack of a slider, the Trombosonic can be played in many different ways. This inspired us to do an informal evaluation to explore the potential applicability of our prototype in different fields and settings other than music. We identified a certain suitability for old and young people and a new possibility for people with restricted mobility to play such a musical instrument. Further development might include a built-in microphone to use the human voice and an expansion of the synthesizer's features.**

Keywords-Sound and Music Computing, Interface for Musical Expression, Exploratory Evaluation

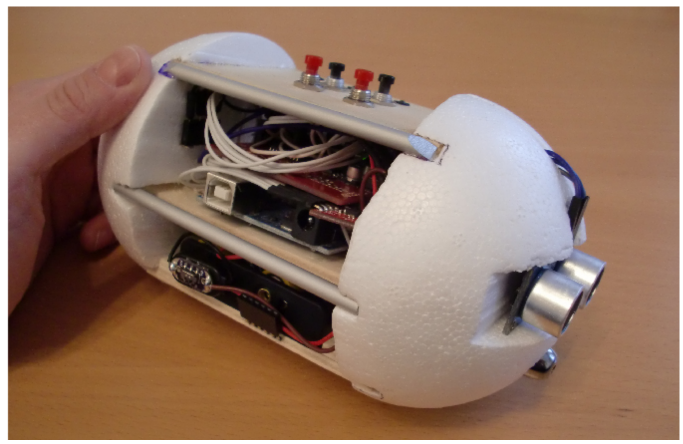


Figure 1: The final Trombosonic prototype

I. Introduction

Musical instruments equipped with sensor-technologies allow many different ways of expression and interaction [1]. Apart from using them for musical purposes the application of such versatile interfaces can be manifold (e.g., [2], [3]).

In this paper, we present a new multi-purpose interface named "Trombosonic". The primary intention was a new digital music instrument inspired by the slide trombone. Hence, we started to design the interface under some self-defined constraints. Unlike many existing approaches, we did not augment a trombone (e.g., [4]) or used the instrument as an example for a digital music interface imitating the trombone's look and feel to create an electronic slide trombone (e.g., [5]). For our development we rather took the technique for playing the trombone as a guiding principle only, to enable an embodied control of sound with a preferably simple and compact hand-held interface.

Our initial design considerations led to preconditions which address sensor-technology and construction. To balance functionality and complexity of the interface and keep it as simple and cheap as possible, we decided to use only standard off-the-shelf low-cost hardware, such as an ultrasonic sensor, push buttons and an accelerometer, to mention some of the important ones. By doing so, we could keep the costs for hardware and material below 100 Euros in total and were still able to explore a range of different sensors within one device.

Throughout the design process, the basic intentions regarding appearance, functionality and materials changed significantly. We initially started with a paper-made tubular prototype to simulate a trombone. The final interface is shown in Fig. 1 and illustrates the visual difference to a traditional slide trombone (see Fig. 2). Most notable is the missing typical slide that characterises a trombone. Despite that, it is played like a trombone with slide motions by holding it in one hand, either left or right, and changing the pitch by moving the other hand back and forth. The name "Trombosonic" is the combination of the two words "trombone" and "sonic".

However, during the development phase it turned out that the device might also be useful for other applications as a hand-held interface. Apart from its original purpose to serve as a musical instrument, an exploratory evaluation has shown its potential applicability in fields such as education, sonification, therapeutic prevention and rehabilitation.

The main contribution of this paper is the presentation of the new digital music instrument Trombosonic along with an exploratory evaluation. We use both expert knowledge and the experience of extreme users [6] to identify potential future applicability in music and non-music domains.

In the following we start with the description of similar research and existing literature our project is built on. Then we go on to describe the design and the functionality of the Trombosonic. Finally, we present the exploratory evaluation

 ISBN: 978-1-61208-325-4

Figure 2: The jazz trombonist Roman Sladek plays a traditional slide trombone (Photographer OhWeh)

Figure 3: Playing the Trombosonic, the red laser in the palm indicates the direction of the ultrasonic sensor

that shows the potential applicability of our prototype.

II. Related work

Both researchers and artists have used the trombone for their work in many ways. Composers appreciate the trombone as "very adaptable system for capturing, suspending and altering shards of sound" using different electronic extensions to create new sounds and sample external sources [7]. Farell augmented a trombone by using a minimal-hardware ultrasonic sensor for the slide, a modified mouthpiece and a loudspeaker in the trombone bell to change the original sound of the trombone for his electro-acoustic performances [4]. A very simple prototype using an optical sensor to detect the position of the slide was created by Lemouton et al. to realise a gestural interface for a traditional trombone [8].

Instead of augmenting existing instruments, Bromwich built a completely new instrument, the Metabone, using only the trombone's dynamics and characteristics [9]. Su et al. present an electronic trombone for the entertainment of children and a playful introduction in musical instruments [10]. The Double Slide Controller derives from the traditional trombone [5]. It looks different though and appears as a complex interface.

Unlike the presented examples that use the trombone as a model or augment an existing instrument, we wanted to combine its most promising features within one simple and compact interface. Keeping in mind the trombone as original instrument and its possibility to create sound by a unique hand gesture, we also wanted to provide new features and embodied interaction that goes beyond the usual musical purposes.

III. Overall concept

The Trombosonic's hand-held interface is purely electronic without any loose or moveable parts. It is held in one hand, either left or right, with a pinch grip. For data processing it uses an Arduino Duemilanove microprocessor [11]. An attached RedFly WiFi-Shield [12] sends sensor data as OSC messages wirelessly to a computer running Max/MSP for sound synthesis in our particular case or any other OSC-compliant musical application.

The casing of the interface is cylindrical with rounded ends and made of polystyrene and wood (see Fig. 1 and Fig. 4). This keeps it lightweight but stable and handy. All electronic devices are bolt-on or glued. Additionally, four aluminium rods provide a good grip and they round out the overall appearance. Its total weight including batteries is 294 g (10.37 oz).

For powering the Arduino, a battery pack is included at the bottom of the interface which holds four standard AA batteries. The longest period of time that the Trombosonic was turned on for testing purposes was 130 minutes and no energy problems were observed during this time. An accurate test regarding energy consumption has not been done yet.

Both the compact design and the wireless network communication ensure free and easy movement during usage within the range of the arm and without being wired to the computer. The whole set of sensors and why they are specifically used to enable embodied musical expression, are described in detail in the following.

IV. Musical expression

The Trombosonic's primary intention is to serve as a musical instrument. Hence, it has several features that enable expressive sound control, including (1) four push buttons and four LEDs, (2) an ultrasonic sensor, (3) a red laser pointer, (4) a thermal resistor, and (5) an accelerometer. For its use as a musical instrument the sensors and actuators are configured to work for particular musical purposes. All features are shown in Fig. 4 as they are located on the device and an overview with a short description is given in Table I.

A. Physical sound generation

Four push buttons, mounted on the top board, enable the control of the basic functions. They are ordered in a square and operated with the middle finger and the ring finger. This allows a good grip using the other two fingers while pushing the four buttons. For additional visual feedback each button is connected to an LED in a different colour, which flashes when the button is pushed. All combinations of how the buttons can be pushed and the corresponding functions are shown in Table II.

The Trombosonic uses a subtractive sound synthesis. The default frequency of the oscillator is 440 Hz. Button 1 turns on the sound, while button 2 allows the user to save and hold the actual frequency which changes continuously according to play. With this function, the player is able to explore the acoustic frequency spectrum endlessly or at least within

 ISBN: 978-1-61208-325-4

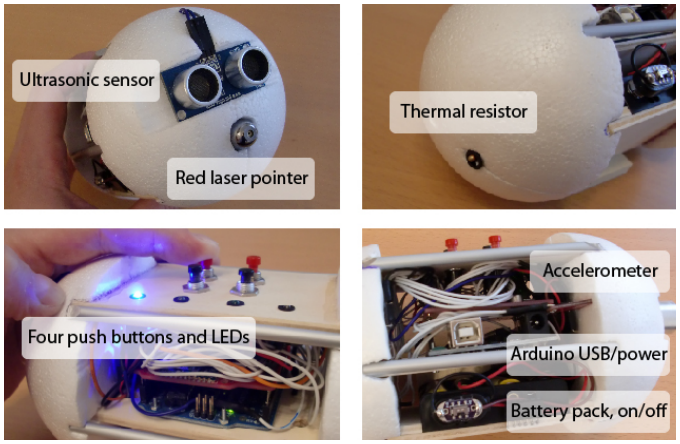


Figure 4: Description of all features of the interface

TABLE I: Overview of features and functionality

Feature	Functionality
Four buttons	Tone on/off, set synthesizer param.
Ultrasonic sensor	Pitch/frequency control
Red laser pointer	Direction of the ultrasonic sensor
Accelerometer	Position/movement of the interface
Thermal resistor	Using the player's breath

the human acoustic range. Buttons 3 and 4 switch between waveforms of the oscillator and a Low Frequency Oscillator (LFO). Pushing buttons 1 and 2 or 3 and 4 together switches between filter types and turns on the laser. Pushing all buttons at once, resets the reference value of the thermal resistor. All functions are described later in detail.

B. *Embodied expression*

The design of Trombosonic enables a range of embodied expressions in play. The **ultrasonic sensor** [13] at the front enables the typical pitch control of the generated tone as known from the slide trombone. Unlike the traditional instrument the Trombosonic has no slider or handle. Instead, the **red laser pointer** (a disassembled off-the-shelf model for presentations) indicates the direction of the ultrasonic sensor for a better orientation of the pitch-steering hand as shown in Fig. 3. While moving it back and forth a red dot is projected on the palm. This realisation allows the player to play the instrument with two hands, comparable to a slide trombone which also makes it familiar to spectators in its embodied movements.

Because the ultrasonic waves can bounce off any object, the second hand is not mandatory, thus the Trombosonic can also be played with just one hand and interact with other objects. These objects may be items within the performer's environment, or the body itself. Whatever interface is pointed on, the distance is transformed into sound. Even spectators who are moving or waving hands can allow interactive sonification of both performer and audience. The laser pointer can also be turned off and on at any time during a performance to avoid dazzling the spectators.

TABLE II: Summary of button functions

Buttons pushed at once	Functionality
1	Tone on/off
2	Set new frequency
3	Switch oscillator wave
4	Switch LFO waveform
1 + 2	Switch filter type
3 + 4	Laser on/off
1 + 2 + 3 + 4	Set thermal resistor value
other button combinations	not used yet

Another embodied sound control is realised with an **accelerometer** [14] that measures the interface's movement in three dimensions. The actual synthesizer implementation uses two of them. The device can be turned around the longest axis (the one the red laser points to) and up- and downwards to control frequencies of the LFO and the filter.

Given that the trombone, the source of our inspiration, is a wind instrument, we also included a mouth piece in our interface. Unlike the slide trombone, it is for additional expression only and not the origin of the tone. For reasons of simplicity we did not use a complex breath analyser [4], [5] but a simple **thermal resistor** [15] to recognise the player's breath. During the design process we used this value to intensify different parameters of the synthesizer, such as the bandwidth of the frequency filter. However, with the actual prototype, the breath control gives the volume a boost as this seems to be comparable with a traditional wind instrument.

C. *Sound synthesis*

For our applications the Trombosonic uses Max/MSP for a subtractive sound synthesis. The whole patch is controlled remotely with the wireless interface and receives nine different sensor-values (see Table I). To generate a sound, the player can choose among simple waveforms which are attenuated by an ADSR (Attack Decay Sustain Release) envelope, an LFO and filter effects. Certain parameters can be controlled in real-time with the Trombosonic's hand-held interface.

For this prototype we focused mainly on the interface and minimised the synthesizer's features. Hence, the sound reminds a little of old synthesizers. Furthermore, there is no special musical training or knowledge needed to play the Trombosonic and to explore its features.

V. Exploratory Evaluation

We did an informal evaluation of the Trombosonic as a musical instrument and explored the potential applicability of our prototype in different fields. For this purpose we asked experts at a competition for new musical instruments and a researcher in game and interaction design.

Additionally we build on knowledge from existing literature about the value of extreme users [6]. The positive impact of music and the suitability of musical instruments in various non-music domains have already been shown. For instance they can be used as playful, toy-like devices for non-specialists (e.g. [3], [16]) or for therapeutic prevention and rehabilitation (e.g. [17], [18]). This inspired us not only to use expert knowledge to evaluate the Trombosonic but also to give it away to people with different abilities and ages such as a 92-year-old woman and a 13-year-old boy. We considered them as untypical users for new musical interfaces and expected them to help to explore the Trombosonic's potential beyond performances.

Everyone participating in the exploratory evaluation was not involved in the project before and saw the device for the first time. After a short introduction they were allowed to play the interface freely. Afterwards they were asked to tell us about their experience. In addition, we took photographs of their explorations and took notes of their comments. All subjects started to play the Trombosonic with its general sonic and gestural features we described in section IV. With each of them we spent about 30-45 minutes for exploration and talked with them about their experience. For some we slightly changed single features tailored to their anticipated interests and needs, as will be described below.

A. New interface for musical expression - expert evaluation

To hand out the Trombosonic to a musician is the most obvious test for a musical instrument. We did so in early spring 2013 for a performance in Vienna where the Trombosonic was used as special instrument for a certain part of a show. The artist used it as a solo instrument during one song.

However, to "fit" better with the other unusual users and to get the most interesting and diverse results, we did something different. We applied for the annual Margaret Guthman Musical Instrument Competition [19] which is considered one of the largest competitions for new musical instruments. The Trombosonic was chosen out of more than 70 submissions to take part in a performance as one of 17 semifinalists which means an acceptance rate of lower than 25%. The successful submission to this highly competitive and renowned competition proves that the Trombosonic is already well-regarded as a new musical instrument. The actual performance took place in Atlanta, USA, in April 2013. We took advantage of this event to get the official feedback of the expert jury as well as the opinion of other participants and audience members when presenting it as a new musical instrument.

The performance at the competition was successful and two pieces were presented: One original electro-acoustic composition and one rather mainstream oriented piece accompanied by pre-recorded playback. People in the audience as well as the jury enjoyed the presentation of the many different features and how the Trombosonic was played in a trombone-like manner during the first piece. The second piece was called "Trombopolka" and was intended to be a tribute to the original instrument, the slide trombone. The Polka is a popular genre of folk music. Some audience members explicitly stated after the performance how they liked the combination of traditional music and the new musical instrument.

The experts mainly criticized deficiencies in the sound synthesis and some spectators missed the acoustic traceability of the breath sensor. One suggestion from another musician was to integrate a microphone for additional sound creation using the human voice. Two other performers pointed out the compact and wireless design which makes it easy for embodied performances as they anticipated.

In summary, the performance at the musical instrument competition confirms the potential of the Trombosonic as a new interface for musical expression and various comments from new musical instrument experts suggest the direction of future revisions and improvements.

B. Physical training for older adults

We then gave the Trombosonic to a 92-year-old woman who is a relative of one of our project members. She was willing to help us for the evaluation during a visit at her own house. She has full mental abilities apart from some forgetfulness from time to time, as she confessed herself. She is still able to walk without a cane in her home. She told us, she uses a walking stick only outside as a precaution and especially during the winter season. However, according to her own description her movement abilities are getting worse and her visibility is already in a bad condition. Asked for her musical knowledge she said, she had learned to play the piano a long time ago and loved to play music and to sing. Now she is unable to play any more since she cannot see the keys and the score.

Figure 5: A 92-year-old woman playing the Trombosonic: First impression (left), standing to operate it "in another way" (right)

We did not present the Trombosonic as a music instrument to her. According to what literature suggests in relation to physical activity and elderly people [20] we rather said it was an acoustic training device. Addressing her own musical experience, we changed the original electronic sound with a piano synthesizer to make it sound more familiar to her. After an explanation of the buttons and some possibilities to make sound, she started to handle it by herself.

Conversation with her and our own observation have shown that the originally intended way to play the Trombosonic with two hands was not very convenient for her. What was notable though, was her behaviour changing her hands holding the device alternately in both hands and finally she even stood up to operate it "in another way" as she noted (see Fig. 5, right). She said she tried to find a good way to hold it and at the same time preventing her arms from getting tired when moving the device by changing hands. Unlike all other participants of the exploratory evaluation, she was the only one considering tiring issues during playing the instrument. This might be important when using the Trombosonic for older adults or rehabilitation.

It appeared to us that she mainly concentrated on the device itself instead of producing particular sounds. However, at the end of our session she summarized her experience: "I really enjoyed making it sound like a piano doing moves I am usually not used to do. Though I do not know how it works and why it sounded like a piano" (Translated from German).

Overall, we identified a certain interest in the Trombosonic and her different ways to handle it. Following Rolland et al. who illustrate that "regular physical activity is a key component of successful aging" [20] and Bruhn and Schröter who discuss the positive impact of making music in old age, we propose the Trombosonic as a potential device for elderly

 ISBN: 978-1-61208-325-4

people. It might be a good way to combine physical and musical activity.

C. Playful interface for children

When talking about musical play and young children, Tarnowski explains "functional musical play might include exploring vocal, instrumental, and environmental sounds as well as the way in which these sounds are made" [21]. This motivated us to give the Trombosonic to a young boy aged thirteen (Fig. 6, right) who was visiting our lab for a trial internship. He has no instrumental training but considers himself a very interested listener to music which is also indicated by the big headphones he wears around his neck all the time. Additionally he started to make music with his computer a little while ago, experimenting with a software-synthesizer.

Similar to the older adult, we explained the basic functionality of the Trombosonic to him and how to handle it. When he started playing we observed, most different to all other evaluation participants, that he really seemed to focus on the music. We also noticed that he played the Trombosonic mostly in its originally intended way using two hands. However, once he started to roll the interface with one hand on the table to create a smooth wave-like sound using the accelerometer. He was the only one who used the movement features of the interface in this physical way together with other objects such as the table.

In all, the young boy carefully analysed the different features and ways to play the Trombosonic throughout his whole session. Following his own words "it was a lot of fun" and he would like to control his own sounds with the interface. We propose the Trombosonic as a suitable instrument for letting young people playfully explore music without being able to play a traditional instrument.

D. Sonification and people with disabilities

Finally, we asked a researcher with expertise in interaction design within our lab to tell us about his experience with the Trombosonic. After an initial explanation of the basic functionality we let him explore the device. It was significant that he started to use it as a one-handed device despite our initial advice to play it in a trombone-like manner. Following his own "intuition" (as he defined it by himself) he started to walk around the room using the Trombosonic as a sonification device. He started to explore the environment acoustically while pointing the device onto different walls and surfaces (Fig. 6, left).

Furthermore, he turned the device around pointing the ultrasonic sensor towards his own body. Moving it back and forth he started explaining: "Look, now it is a one-handed instrument. I can play a trombone without my second hand" (Translated from German). During his test he complained about the lack of clear feedback when using the buttons to control the synthesizer. Since he was not familiar with the synthesizer's options this was really a problem when trying to intentionally switch between wave forms and filters as he said.

The trial with the interaction designer suggests some usability improvements for a more intuitive handling. Furthermore, it might be worth considering the Trombosonic as a one-handed musical instrument keeping in mind that "thousands of people with disabilities in the UK, and millions across the world, are excluded from music making" [22]. The Trombosonic could be such an instrument to enable those

Figure 6: People playing the Trombosonic, as a one-handed instrument acoustically exploring a shelf (left), trying its features played with two hands (right)

people and people with restricted mobility in general to gain a trombone-like musical experience. It could also have potential for people with visual impairment as a way to playfully explore their physical environment.

VI. Discussion

The exploratory evaluation was not meant to be comprehensive but to complement the main contribution of this paper, the presentation of a new musical instrument. It gave us a differentiated impression of how people play the Trombosonic from the perspectives of both experts (in new musical instruments and in interaction design) and extreme users (very old, young). Their considered feedback, as well as their unanticipated uses, pointed to potential applicability that might be worth considering and gave some initial directions for future development.

Overall, people tried various different ways to handle the interface, such as using one-hand only or both hands and while standing, sitting or moving around. The actual usage to produce sound was ranging from playing music following scores in a traditional way to acoustically explore the environment.

The approach of using an exploratory evaluation when testing new musical instruments turned out to be qualified. It was inspiring to use expert knowledge as well as to see unexpected behaviour of unusual users. We argue that our assumption to widen the range for non-obvious applications by doing an exploratory evaluation was verified. At least for the initial of a new musical instrument and a new interaction device this opened a set of unpredictable possibilities for improvements and new directions to focus on for future development.

Compared to the initial described approaches which augment traditional trombones or create new interfaces on the basis of the original instrument, our strategy to create a compact device has its advantages as the exploratory evaluation has shown. Despite its different appearance, people considered the Trombosonic to be a trombone-like instrument. At least when it is played as intended which happened during the instrument competition in our particular case. Analogously people tend to play with the interface in unusual ways and they explore its features as soon as they do not think of it as a trombone-like instrument such as the interaction designer and the older adult.

Thus, designing a new musical instrument under certain constraints and evaluate the prototype in an exploratory manner brought the anticipated insights in unexpected and unpredictable user behaviour. The combination of experts and extreme users helped to go beyond the usual applicability

 ISBN: 978-1-61208-325-4

of this musical interface in fields such as healthcare and education.

VII. Conclusion

The Trombosonic is a new instrument for musical expression that derives from the slide trombone. However, it does not imitate the slide trombone either visually or acoustically, rather the principles of this wind instrument serve as a design inspiration for the interactive gestures.

Push buttons, an ultrasonic sensor and a red laser allow an embodied playing of the instrument similar to the slide trombone changing the pitch with one hand moving back and forth. Compared to a traditional slide trombone, the whole instrument's size is much smaller and the slider is completely missing. Furthermore, an accelerometer and a thermal resistor enable an additional embodied expression. Moving the whole interface enhances the musical possibilities compared to the traditional instrument, while the use of the player's breath retains a typical feature of wind instruments.

Along with presenting the Trombosonic as a new interface for musical expression we did an exploratory evaluation looking for its potential as a musical instrument as well as in other fields. Hence, we successfully submitted a performance proposal to an international competition for new musical instruments and gave the instrument to a 92-year-old woman, a 13-year-old boy and a researcher in game and interaction design. This let us identify different issues and unexpected aspects to keep in mind for future improvements. All cases also indicate the Trombosonic's suitability for various musical purposes as well as non-music applications.

VIII. Future directions

For advanced prototypes of the Trombosonic we plan to integrate a microphone for additional sound creation using the human voice. Furthermore, the synthesizer needs some revision regarding the sound and better mapping of sensor values to single parameters, along with a more intuitive button control. Beyond technical improvement addressing mainly musical features, the evaluation suggests to adapt and use the interface in other domains. It could be used as training device for elderly people addressing physical and musical health-relevant activity or it could let children intuitively explore sound generation without being trained to play a traditional music instrument.

References

[1] C. Dobrian and D. Koppelman, "The'E'in NIME: Musical Expression with New Computer Interfaces," in Proceedings of New Interfaces for Musical Expression, 2006, pp. 277–282.

[2] C. Pacchetti and F. Mancini, "Active music therapy in Parkinson's disease: an integrative method for motor and emotional rehabilitation," vol. 393, 2000, pp. 386–393.

[3] D. Robson, "Play!: Sound toys for non-musicians," Computer Music Journal, vol. 26, no. 3, 2002, pp. 50–61.

[4] N. Farwell, "Adapting the trombone: a suite of electro-acoustic interventions for the piece Rouse," in Proceedings of New Interfaces for Musical Expression, 2006, pp. 358–363.

[5] T. Henriques, "Double Slide Controller," in Proceedings of New Interfaces for Musical Expression, 2009, pp. 260–261.

[6] L. Holmquist, "User-driven innovation in the future applications lab," Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04, 2004, pp. 1091–1092.

[7] N. Collins, "Low Brass: The Evolution of Trombone-Propelled Electronics," Leonardo Music Journal, vol. 1, no. 1, 1991, pp. 41–44.

[8] S. Lemouton, M. Stroppa, and B. Sluchin, "Using the augmented trombone in I will not kiss your f. ing flag," in Proceedings of New Interfaces for Musical Expression, 2006, pp. 304–307.

[9] M. A. Bromwich, "The Metabone: An interactive sensory control mechanism for virtuoso trombone," in Proceedings of International Computer Music Conference, 1997, pp. 2–4.

[10] M. Su, W. Lee, and S. Chen, "Electronic trombone: an interactive tool to promote musical learning and performance creativity," in Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry, 2011, pp. 585–588.

[11] Arduino. [Online]. Available: http://www.arduino.cc (last access 24. January 2014)

[12] RedFly WiFi-Shield. [Online]. Available: http://www.watterott.com/de/Arduino-RedFly-Shield (last access 24. January 2014)

[13] Ultrasonic sensor. [Online]. Available: http://www.watterott.com/en/Parallax-PING-Ultrasonic-Sensor-28015 (last access 24. January 2014)

[14] Accelerometer ADXL335. [Online]. Available: http://www.watterott.com/en/Breakout-Board-ADXL335 (last access 24. January 2014)

[15] Thermal resistor. [Online]. Available: https://www.sparkfun.com/products/250 (last access 24. January 2014)

[16] S. Jordà, "Digital Instruments and Players: Part I Efficiency and Apprenticeship," in Proceedings of New Interfaces for Musical Expression, 2004, pp. 59–63.

[17] C. Plahl, "Musiktherapie - Praxisfelder und Vorgehensweisen," in Musikpsychologie. Enzyklopädie, Rowohlts, 2009, pp. 630–652.

[18] M. E. Clark, A. W. Lipe, and M. Bilbrey, "Use of music to decrease aggressive behaviors in people with dementia." Journal of gerontological nursing, vol. 24, no. 7, Jul. 1998, pp. 10–17.

[19] Margaret Guthman Musical Instrument Competition. [Online]. Available: http://www.gtcmt.gatech.edu/guthman2013 (last access 24. January 2014)

[20] Y. Rolland, G. Abellan van Kan, and B. Vellas, "Physical activity and Alzheimer's disease: from prevention to therapeutic perspectives." Journal of the American Medical Directors Association, vol. 9, no. 6, Jul. 2008, pp. 390–405.

[21] S. M. Tarnowski, "Musical Play and Young Children," Music Educators Journal, vol. 86, no. 1, 1999, pp. 26–29.

[22] The One-Handed Musical Instrument Trust. [Online]. Available: http://www.ohmi.org.uk (last access 24. January 2014)

Colourful Privacy

Designing visible privacy settings with teenage hospital patients

Maja van der Velden
Department of Informatics
University of Oslo
Oslo, Norway
majava@ifi.uio.no

Margaret Machniak
Department of Informatics
University of Oslo
Oslo, Norway
margarma@ifi.uio.no

***Abstract*—The paper reports from a qualitative study based on the analysis of semi-structured interviews and Participatory Design activities with hospitalised teenagers with chronic health challenges. We studied how teenage patients manage their online privacy, with a focus on the design and use of privacy settings. We found that the majority of participants preferred to visualise privacy settings through the use colours and to personalise access control. They also considered these necessary on more secure patient-centred social media. As proof of concept, we implemented some of the findings in a patient social network setting. We conclude that visualising and personalising privacy settings enable young patients to have more control over the sharing of personal information and may result in a more effective use of privacy settings. In addition, privacy-aware default settings may prevent teens from unintended sharing of personal information.**

Keywords-Facebook; participatory design; patient social media; privacy settings; teenage patients; visualisation of privacy

I. Introduction

From a developmental perspective, the teenage years (12-18 years old) are characterised by a quest for identity, independence, and autonomy [1]. New digital technologies, such as social media and mobile phones, are described as supporting this quest [2], [3]. These technologies are therefore now also explored as platforms to provide support to teenagers who are dealing with chronic health challenges, such as diabetes, asthma, cancer, chronic organ diseases, rare diseases, etc. There is a growing interest in the design and development of new applications, and several initiatives actively involve young patients in the design process [4]–[9].

Chronically ill teens are born with a medical condition or they develop such a condition during their childhood or their teenage years. More and more children born with chronic diseases survive into their adult years [10]. As a result, the teenage years are seen as a transition period, in which teens begin to take more responsibility for the treatment and other activities required by their diagnosis [11]. They may therefore have additional information and communication needs and interests because of their chronic diagnosis and treatment.

In Participatory Design, new interactive technologies are not only designed for young patients – they are designed with them [9], [12], [13]. Designing with young patients and other vulnerable users is perceived as difficult as additional ethical issues may arise because of particular circumstances [14], [15], or it may be difficult to include teenage patients because of self-esteem or identity issues [16]–[18].

In our study, we met teenage patients while they were receiving treatment in the hospital. This decision was based on our earlier experiences in research with teenage patients. In a 2011 field study, we looked among other things at the use of a closed social network for young patients in Canada [18], which was often only accessible from a computer inside the hospital. An user account could only be obtained from hospital staff. In order to meet some of the users, we needed to implement the study in one of the participating hospitals.

We were especially interested in understanding how teenage patients *practice* privacy on social media. The practice of privacy can be understood as an expression of their autonomy [19]. We found that many teens don’t discuss their diagnosis and treatment with their friends and followers on social media. The reason for this was that they often tried to be “normal” or “regular” teens [18], [20]. Being chronically ill wasn’t something they would discuss in social media or sometimes not even outside the physical surroundings of the hospital or the security of the direct family:

“*I guess I just pretend I am normal and I don't have it when I am outside the hospital*.” (girl, 17 yrs.)

“*I like to be as regular as I can, so I don't want to talk about it.*” (boy, 13 yrs.)

We explained this as a form of self-protection and self-presentation [19]. For the outside world, the teens separate their identity as patient from their identity as teenager. Thus, we decided to implement the new study again in the Children’s Hospital of Eastern Ontario (CHEO). In order to meet the teens as patients, we needed to meet them where they felt secure to talk from a patient perspective.

This study is part of a larger research and design project called KULU [21]. The objective of KULU is to support young patients (12 – 25 yrs.) with chronic health challenges in their autonomy. One of the KULU projects is the design of a closed social network for teenagers with chronic health challenges. In this paper we present and analyse the design of privacy settings for such a social network. The rest of the paper is as follows: In the next section, we briefly discuss some of the research done on this topic. In Section III, we describe our methodology, methods, and some of the challenges we encountered. In Section IV, we present our findings, which we discuss in Section V. In Section VI, we implement some of the findings in a prototype for a closed social network for teenage patients. In the final section we present our conclusion and future work.

II. BACKGROUND

The use of personal health information is regulated by national policies and guidelines in order to prevent inappropriate use. Social media and other online sites seriously challenge this regulation: patients voluntarily share their health information online to gain advice and support, or to support others [22], while some medical personnel post images of patients or discuss particular cases with colleagues [23]. Insufficient use of a system's privacy settings, or the copying of information to a site with a less restrictive privacy policy, may unwantedly disclose sensitive health information. This may affect the patient's access to work, insurance, and relationships [24].

A. *Design of privacy settings*

There is a large body of literature reporting on privacy management, privacy practices, and privacy settings on social networking sites, e.g., [25]–[28], but only a small portion has the design of privacy settings as its focus [29]–[33]. Privacy settings on social networks are the customisable options and technologies that regulate the accessibility of personal data to other users and to third parties. In this context, privacy can be defined as "a set of practices in negotiating the public and private divide" [29]. The design of privacy settings of social networks thus refers to the design of options and technologies that enable users to negotiate this divide. Existing patient social networks, such as Upopolis, PatientsLikeMe, and Mayo Clinic Connect don't use (customisable) privacy settings.

B. *Visualising privacy through use of colours*

Only a few studies address the use of colours to visualise privacy levels to users of Internet services [34]–[36]. PrivacyDefender, now discontinued, introduced as a Facebook app in 2010, helped Facebook users to understand their privacy settings, but is no longer available [34]. C4PS, now discontinued, was a Firefox browser plug-in using colours to help represent privacy settings on Facebook [32].

III. METHODOLOGY

This study takes a qualitative approach towards the study of teenage patients, social media, and privacy and is guided by the following questions:

- Can design interventions support teenage patients in managing their privacy settings?
- Does a closed social network for young patients need privacy settings?

A. *Participatory Design*

This study is based on Participatory Design principles [37], [38]:

- Those who will use the technology should have a voice in the design
- Mutual learning between users and designers based on their respective skills, interests, values
- Starting the design process in the practice of the users

We build forth on the teens' existing social networking experiences and skills by using examples from Facebook.

B. *Ethical challenges*

In the study we adhered to the research ethics and privacy requirements stipulated by the Norwegian Data Inspectorate and the Research Ethical Board of CHEO. There were two additional ethical challenges we needed to deal with, on-going consent and parents. On-going consent is especially important in situations in which the participant is not able to leave the research setting without help. Secondly, most of the teens were accompanied by a parent or other adult relative. The Youth Council of CHEO reminded us to ask the teens if they wanted their parent(s) present or not.

C. *Logistical challenges*

The particular way in which hospital patients were recruited for the study had an important effect on the design of the study. It was known from the onset that we could not count on more than 30-40 minutes contact time with a patient and that we would meet each teen individually. The majority of meetings took place in a treatment room in the polyclinic or a patient room on one of the wards. In all these cases, the teens were lying in bed and the majority was receiving intravenous treatment while participating in the study. This means that they weren't mobile and had minimal use of one arm.

The rooms were furnished with hospital tables, which could be used for design activities. The size of the tables was 40 x 80 cm and could be installed over the bed. The tables had to be used for the audio equipment of the researcher, to record the session, and storage and use of design materials. The table had to be movable at all times, in order to make room for hospital staff. The design materials themselves needed to be made of materials that could be disinfected when they were used by different patients.

TABLE I. SEMI-STRUCTURED INTERVIEW

Interview themes
1. Technology used in the hospital (mobile phone, laptop, etc.) 2. Favourite sites or things to do on the Internet 3. Age when starting to use Facebook 4. Do you know your privacy settings? 5. Does a closed social network for young patients need privacy settings?

D. *Methods*

The limited contact time, small workspace, and the need to be flexible in terms of being able to pack things up quickly, if needed, affected our choice of methods. We used semi-structured questions and re-design activities based on paper prototypes as our main methods:

1) Semi-structured interviews

The interview was based on a set of questions that would support the analysis of the paper prototyping (see Table I).

2) Paper prototyping

We based the prototyping activities on the example of an existing social network site (Facebook). The first activity was the comparison of three options for locating privacy settings. The first one was the existing situation and the two others presented possible new situations. The comparison was presented as a game (see Fig. 1), which was used to elicit a discussion about the visibility of privacy settings.

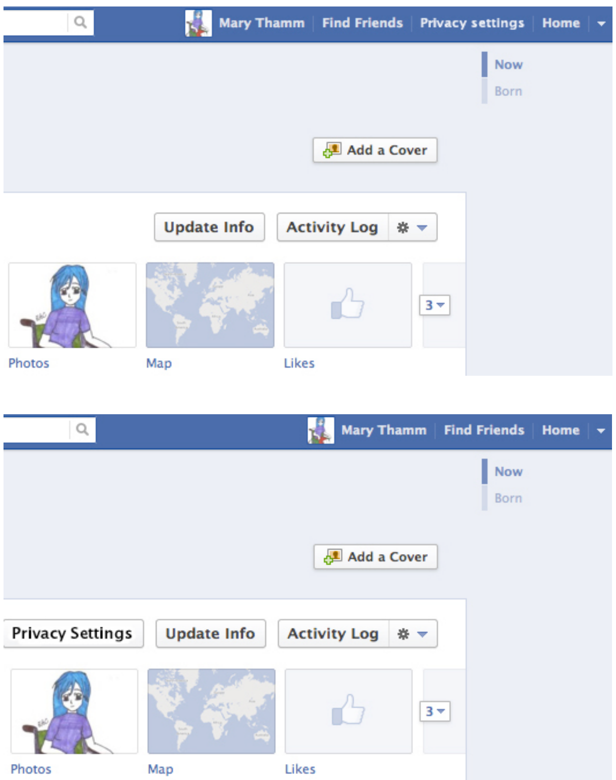


Figure 1. Spot the difference

The second activity was prototyping privacy settings, with a focus on access control: *Who on Facebook can see your status update?* The participants were introduced to using colours to make privacy settings visible. They were given a box with 12 colours and a paper-based image of a status update window. The first image showed the status update window in Facebook. The second image showed the status update window in which the privacy settings were made visible through colours and through self-assigned levels of access. The colours were based on the colour scheme designed by [32]. The participants could use the same colours or personalise the colour scheme.

This prototyping exercise was used to investigate if colours and groups could support the management of their privacy settings. The teens could use Facebook's existing accessibility settings (e.g., 'friends-only', 'friends-of-friends' etc.) or customise these settings so they would better fit their own preferences. We also discussed the need for privacy settings for a closed social network for young patients.

E. Recruitment and participants

The recruitment of the teenage patients participating in the study was undertaken by the Child Life Specialists of CHEO. They knew the teens personally or they could consult their doctor or nurse. Teens were recruited from the wards and from the medical day clinic. They needed to be well enough to participate in a 30 to 40 minute meeting with the researcher. They would receive a leaflet with information and a consent form. The leaflet mentioned that the interview is anonymous, that the teen's diagnosis is not a topic in the interview, and that the researcher has no access to the teen's medical file. Both participant and researcher would sign the consent form. If the teen was younger than 16 years old, a parent also had to sign the consent form.

F. Sample size

In qualitative research, sample sizes differ widely. It is perceived as acceptable to have a small sample size, between 4 and 10 in studies with participants with impairments or other vulnerabilities [39], [40]. In this study, sixteen teenagers participated in the study: nine girls and seven boys. Their ages were between 14 and 18 years old. They participated in the study in the autumn of 2012, while receiving treatment in the hospital.

IV. RESULTS

A. Technology and Internet use in the hospital

All participants brought their personal technologies to the hospital. In addition, some patients used equipment provided by the hospital (see Table II).

TABLE II. TECHNOLOGIES IN USE

Technology	Owned by patients	Owned by CHEO	Used for Internet
Mobile phone	16		15
Laptop	6	2	2
MP3/4 player	6		1
Tablet	1		1
Video game	1	2	
Computer		1	1
DVD player	1		

Facebook and Twitter were the most popular social media. All 16 teens were active on Facebook and nine had a Twitter account. Fifteen out of the 16 Facebook users used their smart phones to receive Facebook notifications and to check their Facebook account while in the hospital.

In order to start a Facebook account, the user needs to confirm that s/he is 13 years old or older. Eight of the 16 teens answered they lied about their age when they registered for their Facebook account. Four of them said they were 13

 ISBN: 978-1-61208-325-4

years and four teens said they were adults, i.e. over 18 years old, while they were all 11 or 12 years old.

B. Privacy settings

Most teens changed their privacy settings only after they started using Facebook. The majority has *Friends-only* as privacy settings (11), while 3 teens have *Friends-of-friends* and one teen had *Public*. One teen had closed down her Facebook account, and was using her boyfriend's account.

Many of the teens were concerned with having the right privacy settings. They would check regularly or find had other tactics to make sure they knew what they were:

"*I've set them up, but I've heard you can make it much more complicated and complex and I read this article about it and it seems very difficult and hard to understand. I don't really find it easy to use [them]*" (girl, 14 yrs.)

"*It is not difficult to get your privacy settings the way you like*" (girl, 16 yrs.)

"*I only post if it is for everyone to see and I just keep everything I post on Facebook to a minimum so if everyone sees it it's not like I'm showing like myself to the entire world*" (girl, 14 yrs.)

C. Making privacy visible

Many teens needed some time to spot the difference between the real situation and the two designed options for locating privacy settings. All but one participant preferred a more visible privacy settings button on all Facebook pages:

"*I already have my privacy settings done, but if I was joining a new social networking program and it had that button, I would definitely click on it*" (boy, 17 yrs.)

"*Because it makes people more aware of their privacy and makes them take initiative I guess you could say instead of it being an afterthought*" (boy, 18 yrs.)

In the case of prototyping privacy settings, all but one participant wished to use colours to make their privacy settings more visible and to dedicate one colour to one particular group of Facebook contacts. One teen was satisfied with how Facebook provided privacy settings for status updates.

Fifteen teens chose to customise their privacy settings: "*That would be cool to set them yourself. [...]. I really do like the idea of personalising those colours*" (girl, 17 yrs.). They created between 2 and 5 categories to organise their Facebook contacts: two participants created 2 categories; five participants created 3 categories; seven participants created 4 categories, and one participant created 5 categories.

Fourteen out of 15 teens created a *Public* category and two teens created an *Only me* category. Most diversity was found in organizing friends. Eight teens created a *Close friends* category and eleven teens created a *Friends* category. Five teens had a *Family* category while one teen created a *Friends without family* and a *Friends with family* category. Five teens created a *Custom* category. This category allows the user to select names from the list with all friends. This customised list would change, depending on the content of the message (status update, photo, etc.) (see Table III). None of the teens used the colour scheme of the example; all preferred to personalise the colour scheme. The most popular colours were the same colours as used in the example, but they were used in combination with other colours.

TABLE III. CATEGORIES

Participants	Categories
14	Public
11	Friends
9	Close friends
5	Family
4	Custom
4	Friends of friends
2	Only me
1	Friends and family

D. Are privacy settings needed?

Eleven of the sixteen participants were asked about the need for privacy settings in a closed social network for teenage patients. Ten mentioned that such settings were needed, "*Oh, of course, there are some things you shouldn't share obviously*" (boy, 18 yrs.).

V. DISCUSSION

Studies have shown that most teenagers with chronic health challenges don't like talking about their experiences outside their direct family and best friends environment [18], [20]. It isn't cool to be different. Teenage patients are careful sharing their personal health information in social media [18]. This was not different with this group of patients:

"*Last time I was in the hospital, I really didn't want anyone to know what was going on because I didn't want rumours to spread*" (girl, 17 yrs.)

The participants in the study were all interested in more immediate access to privacy settings (a large button) and simpler and transparent privacy settings. Several of the teens were not able to locate the privacy settings on Facebook, which (at that time) only became visible after selecting a very small triangle at the top of the page.

Using colours to have an immediate and visible indication of one's privacy setting before sharing a status update was perceived as very helpful and easy to understand. The chosen colour scheme could be used for a variety of purposes, such as visualizing access to one's profile information, photos, etc. Many teens mentioned that they would like the option to personalise their Facebook page, which is not possible at the moment. Choosing one's own favourite colours for privacy settings was therefore perceived as *cool*. Many of them wanted the same colour scheme for use in a closed patient social network.

Although all 15 colour schemes were different, red, yellow, green, and blue were the most used colours. Interestingly, the colours red (5x) and green (6x) were both used to indicate the *Public* setting. In [32], which builds forth on the traffic light colour combination, the colour green is used to indicate *Public*. Some of the participants in our study indicated that green meant 'safe', i.e. the opposite of *Public*: you know exactly who can read your posting. Some of the participants, who used the colour red for *Public*, understood this colour to mean 'attention': "*I kind of see red as, eh, it's something you should be a little more careful about*" (g/16).

There is a clear need for privacy-friendly default settings. Many of the teens lied about their age when they first started using Facebook. New Facebook users have to confirm that they are 13 years or older, and then provide their date of birth in the account settings. Users who were 13-18 years old, had *Friends-of-friends* as default settings and could changed this only into *Friends-only*. The Public setting was not available to this age group (this has changed in 2013). Those who used an age over 17 ended up with *Public* as default privacy settings. In addition, several of the participants have made mistakes on Facebook, sending a status update or photo to the wrong person(s).

A. *The meaning of colours*

Red, green, blue, and yellow, the four most popular colours for privacy settings, are chromatically unique colours and universal in terms of their colour terms [41], [42], but the associations with these colours differ widely [43]. The traffic light colour scheme (red, yellow, green) may have gained universal meaning in the context of traffic, but fails to do so in other contexts. The study made clear the contradictive meaning of the colours red and green in the context of the design of privacy. Colours should thus always be accompanied with a textual indication of their meaning in the particular context.

B. *Limitations of the study*

The particular context of this study, teenage patients in an hospital setting, marks the findings. The study offers meaningful insight into the privacy understandings and needs of this particular group of social media users.

VI. Proof of Concept

We have implemented the findings of the study in one of our prototypes of a closed social network for teenagers with health challenges. Only three categories and three colours were used: *Public, i.e. all users of the closed social network,* (green), *Friends* (yellow), and *Only me* (red), see Fig. 2. The default privacy setting for all users is *Only me*. Users can configure these settings by selecting the large 'Settings button', visible on every page (see Fig. 2). The default setting for all status updates, photos, etc., always remained *Only me*, forcing the user to make a choice every time something new is shared with other users. The different settings can be applied to all types of content on the site as a response to higher privacy demands on a closed social network as opposed to Facebook.

The prototype was evaluated by a group of teenagers in Norway, patients and non-patients. The teens appreciated the possibility to alter the privacy settings for each new piece of information. They especially liked the combination of overall settings, which could be altered using the 'Settings button', and the flexibility to choose the audience for each post in situ.

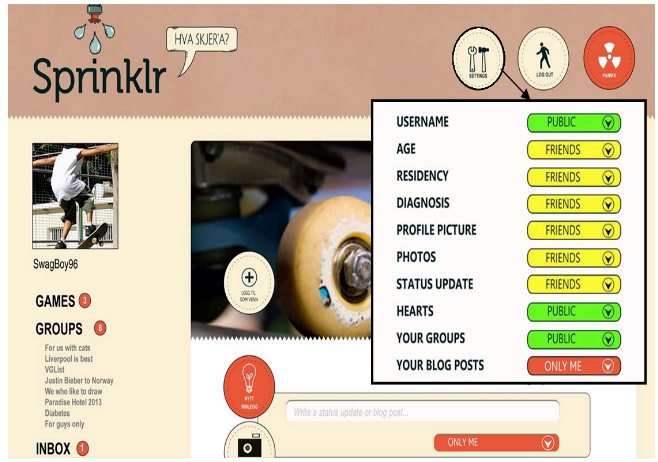


Figure 2. Proof of concept

The teens appreciated that all default settings were set to *Only me*. They elaborated that this was *cool* and gave them the feeling of control: they didn't have to share anything about themselves if they didn't feel like it. The participants thought that the choice of colour was appropriate, as the red colour indicating the *Only me* setting was more prominent and drew their attention to it. This applied only because the default setting was *Only me* and had to be changed in order for others to see the post. The teens also suggested that the *Only me* settings could be used to expand the functionality of the site – the posts marked as *Only me* could for example function as a diary.

VII. Conclusion and Future Work

The combination of the participants' Facebook experiences and skills and the researchers' design skills resulted in privacy-aware designs for a closed social network for young patients:

- A large button for configuring privacy setting
- Privacy-aware default setting for sharing new information (status update, posting, photo)
- Use of colours to visualise privacy settings

Personalisation, being able to choose one's own colours to visualise privacy settings, may motivate teens to use privacy settings more effectively.

In our on-going research on the design of privacy settings, we focus on the use and meaning of colours.

Acknowledgement

Maja gratefully acknowledges the support of the Child Life Specialists of the Children's Hospital of Eastern Ontario (CHEO) and Dr. Khaled El Emam, head of the Electronic Health Information Lab (EHIL) at the CHEO Research Institute, during the implementation of the study. We thank all participating teenagers for sharing their experiences and creativity. The study is funded with a research grant from the Norwegian Research Council.

References

[1] L. Steinberg and A. S. Morris, "Adolescent Development," J. Cogn. Educ. Psychol., vol. 2, no. 1, Jan. 2001, pp. 55–87.

[2] C. Davies and R. Eynon, Teenagers and Technology. Hoboken: Taylor and Francis, 2013.

[3] S. Livingstone, "Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression," New Media Soc., vol. 10, no. 3, Jun. 2008, pp. 393–411.

[4] M. J. Allsop, R. J. Holt, and M. C. Levesley, "The engagement of children with disabilities in health-related technology design processes: Identifying methodology," Disabil. Rehabil. Assist. Technol., vol. 5, no. 1, 2010, pp. 1–13.

[5] L. S. Franck and G. Noble, "Here's an idea: ask the users! Young people's views on navigation, design and content of a health information website," J. Child Health Care, vol. 11, no. 4, Dec. 2007, pp. 287–297.

[6] J C. M. Ruland, J. Starren, and T. M. Vatne, "Participatory design with children in the development of a support system for patient-centered care in pediatric oncology," *J. Biomed. Inform.*, vol. 41, no. 4, Aug. 2008, pp. 624–635.

[7] A. R. Lang, J. L. Martin, S. Sharples, J. A. Crowe, and E. Murphy, "Not a minor problem: involving adolescents in medical device design research," Theor. Issues Ergon. Sci., 2012, pp. 1–12.

[8] E. Mazzone, J. C. Read, and R. Beale, "Design with and for disaffected teenagers," in Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges, New York, NY, USA, 2008, pp. 290–297.

[9] A. Waller, V. Franklin, C. Pagliari, and S. Greene, "Participatory design of a text message scheduling system to support young people with diabetes," Health Informatics J., vol. 12, no. 4, Dec. 2006, pp. 304–318.

[10] S. E. McLaughlin et al., "Improving Transition From Pediatric to Adult Cystic Fibrosis Care: Lessons From a National Survey of Current Practices," Pediatrics, vol. 121, no. 5, May 2008, pp.1160-l166.

[11] S. T. Callahan, R. F. Winitzer, and P. Keenan, "Transition from pediatric to adult-oriented health care: a challenge for patients with chronic disease," Curr. Opin. Pediatr., vol. 13, no. 4, 2001, p. 310.

[12] D. M. Berwick, "What 'Patient-Centered' Should Mean: Confessions Of An Extremist," Health Aff. (Millwood), vol. 28, no. 4, Jul. 2009, pp. w555–w565.

[13] C. I. Reis, C. S. Freire, J. Fernández, and J. M. Monguet, "Patient Centered Design: Challenges and Lessons Learned from Working with Health Professionals and Schizophrenic Patients in e-Therapy Contexts," in ENTERprise Information Systems, M. M. Cruz-Cunha, J. Varajão, P. Powell, and R. Martinho, Eds., Berlin, 2011, pp. 1–10.

[14] A. R. Lang, J. L. Martin, S. Sharples, J. A. Crowe, and E. Murphy, "Not a minor problem: involving adolescents in medical device design research," Theor. Issues Ergon. Sci., 2012, pp. 1–12.

[15] A. L. Culén and M. van der Velden, "The Digital Life of Vulnerable Users: Designing with Children, Patients, and Elderly," in Nordic Contributions in IS Research, M. Aanestad and T. Bratteteig, Eds. Springer Berlin Heidelberg, 2013, pp. 53–71.

[16] P. Chira et al., "Living Profiles: Design of a health media platform for teens with special healthcare needs," J. Biomed. Inform., vol. 43, no. 5, Supplement, Oct. 2010, pp. S9–S12.

[17] R. Mack, E. Giarelli, and B. A. Bernhardt, "The Adolescent Research Participant: Strategies for Productive and Ethical Interviewing," J. Pediatr. Nurs., vol. 24, no. 6, Dec. 2009, pp. 448–457.

[18] M. van der Velden and K. El Emam, "'Not all my friends need to know': a qualitative study of teenage patients, privacy, and social media," JAMIA, vol. 20, no. 1, Jan. 2013, pp. 16–24.

[19] M. van der Velden, "In/Visible Bodies. On patients and privacy in a networked world," in Cultural Attitudes Towards Communication and Technology 2012, M. Strano, H. Hrachovec, F. Sudweeks, and C. Ess, Eds. Murdoch: Murdoch University, 2012, pp. 199–211.

[20] M. A. Applebaum, E. F. Lawson, and E. von Scheven, "Perception of transition readiness and preferences for use of technology in transition programs: teens' ideas for the future," Int. J. Adolesc. Med. Health, vol. 25, no. 2, Jan. 2013, pp. 119-125.

[21] "KULU" [Online]. Available: http://www.kulu.no/english.php. [Accessed: 07-Jan-2014].

[22] S. Fox, "Social Life of Health Information," Pew Research Center [Online]. 2011. Available http://www.pewinternet.org/Reports /2011/Social-Life-of-Health-Info.aspx [Accessed: 07-Jan-2014].

[23] T. Lagu, E. J. Kaufman, D. A. Asch, and K. Armstrong, "Content of Weblogs Written by Health Professionals," J. Gen. Intern. Med., vol. 23, no. 10, Oct. 2008, pp. 1642–1646.

[24] C. L. Anderson and R. Agarwal, "The Digitization of Healthcare: Boundary Risks, Emotion, and Consumer Willingness to Disclose Personal Health Information," Inf. Syst. Res., Apr. 2011, pp. 469-490.

[25] M. Anwar, P. W. L. Fong, X.-D. Yang, and H. Hamilton, "Visualizing Privacy Implications of Access Control Policies in Social Network Systems," in Data Privacy Management and Autonomous Spontaneous Security, J. Garcia-Alfaro, G. Navarro-Arribas, N. Cuppens-Boulahia, and Y. Roudier, Eds. Springer Berlin Heidelberg, 2010, pp. 106–120.

[26] M. Madejski, M. Johnson, and S. M. Bellovin, "The Failure of Online Social Network Privacy Settings," Columbia University Computer Science Technical Reports, 2011.

[27] S. Z. Ibrahim, A. Blandford, and N. Bianchi-Berthouze, "Privacy Settings on Facebook: Their Roles and Importance," in IEEE International Conference on Green Computing and Communications, 2012, pp. 426–433.

[28] M. Netter, M. Riesner, M. Weber, and G. Pernul, "Privacy Settings in Online Social Networks – Preferences, Perception, and Reality," in 2013 46th Hawaii International Conference on System Sciences (HICSS), 2013, pp. 3219–3228.

[29] S. F. Gürses, R. Rizk, and O. Günther, "Privacy Design in Online Social Networks: Learning from Privacy Breaches and Community Feedback," in ICIS, 2008, p. 90.

[30] J. Ausloos, "Guidelines for Privacy-Friendly Default Settings," 20-Feb-2013. [Online]. Available: https://lirias.kuleuven.be/handle /123456789/416312. [Accessed: 11-Nov-2013].

[31] P. Kaur, "Designing User Centric Online Privacy," in Aalto University, Seminar on Network Security, Fall, 2011, vol. 4, p. 83.

[32] T. Paul, M. Stopczynski, D. Puscher, M. Volkamer, and T. Strufe, "C4PS - Helping Facebookers Manage Their Privacy Settings," in Social Informatics, K. Aberer et al., Eds. Berlin: Springer, 2012, pp. 188–201.

[33] H. R. Lipford, G. Hull, C. Latulipe, A. Besmer, and J. Watson, "Visible Flows: Contextual Integrity and the Design of Privacy Mechanisms on Social Network Sites," in International Conference on Computational Science and Engineering, 2009. CSE '09, vol. 4, 2009, pp. 985–989.

[34] "Facebook Privacy Settings Made Easy with PrivacyDefender," Walyou. [Online]. Available: http://walyou.com/facebook-privacy-settings-privacydefender/. [Accessed: 21-Nov-2013].

[35] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder, "A 'nutrition label' for privacy," in Proceedings of the 5th Symposium on Usable Privacy and Security - SOUPS '09, Mountain View, California, 2009.

[36] Web of Trust, "Safe Browsing Tool | WOT." [Online]. Available: http://www.mywot.com/. [Accessed: 21-Nov-2013].

[37] J. M. Greenbaum and M. Kyng, Design at work: cooperative design of computer systems. Hillsdale, N.J.: L. Erlbaum Associates, 1991.

[38] J. Simonsen and T. Robertson, Routledge international handbook of participatory design. New York: Routledge, 2013.

[39] D. J. Lazar, D. J. H. Feng, and D. H. Hochheiser, Research Methods in Human-Computer Interaction. John Wiley & Sons, 2010.

[40] I. Holloway and S. Wheeler, Qualitative Research in Nursing and Healthcare, 3rd ed. Oxford: John Wiley and Sons, 2010.

[41] T. Regier, P. Kay, and R. S. Cook, "Focal colors are universal after all," Proc. Natl. Acad. Sci. U. S. A., vol. 102, no. 23, Jun. 2005, pp. 8386–8391.

[42] F. Ratliff, "On the Psychophysiological Bases of Universal Color Terms," Proc. Am. Philos. Soc., vol. 120, no. 5, Oct. 1976, pp. 311–330.

[43] P. Russo and S. Boor, "How Fluent is Your Interface?: Designing for International Users," in Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems, New York, NY, USA, 1993, pp. 342–347.

On the Mesurement of Mental Models for Interface Design

Ryota Mori
Department of Life Science
Nagano Prefectural College
Nagano City, Japan
e-mail: mori@nagano-kentan.ac.jp

Toshiki Yamaoka
Department of Design and Information Science
Faculty of Systems Engineering
Wakayama University
Wakayama City, Japan
e-mail: yamaoka@sys.wakayama-u.ac.jp

***Abstract*—The purpose of this paper is to provide a measurement of user's thought process after discussing the experiment with "degree of formed mental model", which involves a logical thinking. We studied 42 people based on questionnaire to measurement mental models. This method helped understanding the potential effectiveness of mental model measurement.**

Keywords-mental model; interface; cognition.

I. INTRODUCTION

We describe a method for measuring already formed mental models. The cognitive properties of people and interface designing for them are discussed in this paper.

An interesting approach is presented in this research. We would like you to pay attention to the operating side of the mental model system. This research is peculiar from the standpoint of supporting the operating side as a necessity in the constitution of a society [1].

It is necessary to apply the new method in the solution of usability problems. These problems in the system depend on mental model for 80 % [2]. Mental model is the dominant factor that determines the properties of usability. Mental model varies significantly according to cues on the interface. It is important for user that the amounts of information of user interface design on product is reduced.

It is assumed that mental model approach plays an important in an interaction between the users and the machines. When investigating accidents that happen in socio-technical systems, such as a lack of consideration of human factors generally occurs at the stage of designing the system and its operation. Mental model systems can cause accidents by two or more factors (the complexity and the unexpected error of the system) coming in succession. Therefore, micro-ergonomics [3] exist as a methodology to solve various problems concerning systems and the people who inhabit them.

The definition of mental model are described as follows. Mental model is an eagerness and willingness to do something or the reason why a people wants to do something. To determine motivation, the motivation and guarantee factors are needed. But, the guarantee factor is not referred to our research.

The mental model system refers to the set, the organization, the system, and the mechanism of elements for an organic relation of two or more elements affecting each other, and the settlement as a whole to function.

Many studies have been conducted on the correlation between mental model and cognition [4][5]. Many conventional studies have discussed the results of game or puzzle, used with logic [6][7].

One of the oldest subjects still discussed now is "what is the thinking process on products?" [5]. However, the past studies have not answered the question of their thought. A user interface evaluation has not been performed in detail with the proper amounts of cues for users. Few systematic studies have been reported on the user's thinking.

In general, designing by the use of labeling as a cue on the products is a failure [7]. But, when operating a machine, we have used label by the buttons as an important cue. We could predict the action of machine by the use of labels on the interface.

Usability testing is the most effective evaluation through which one can know if the designer provided the right model to users [2]. But then, usability testing is costly, and the user's stress and cognitive load increases with an increase in testing time. Usually, we do not know about users formed mental models and how user's mental models matched designer's models.

It is important that designer's model is consistent with user's model. A high degree of formed mental model means that user formed a correct model. We have to understand human as a component of the system; in a word, a participant. Therefore, it is necessary to measure the people's mental model to understand the internal structure of the system. How to help a user to form a correct model is provide cues such as affordability, mapping, and straight for the user.

Formed mental model is determined by 9 factors [4]. Questionnaires based on these factors have been designed; the degree of formed mental models is identified using these questionnaires [5].

However, this approach cannot explain how the users' mental models were built, until a user interacts with a product. Moreover, to clarify it, getting interaction interpretation takes some time. To avoid these problems, we have devised a new method by using a quantitative evaluation.

The purpose of this paper is to discuss an approach which quickly evaluates the degree of formed mental model and the interface design suitable for peoples' mental model, while considering the cognitive properties of a user when using an interface. Aside from helping the reader in the pencil - and - paper solution of problems, the analytical skill, which this paper aims at developing, may be useful in special situations. This paper will not go into detailed logic analysis, presuming the reader is familiar the reader will be familiar with the people's memory and thought perception. We discuss a situation which assumes the people are not very familiar with the control procedure of electrical appliances.

We have to comprehend the circulation of information. To do so, it is necessary to understand the relation of elements that prevent forming a mental model. Therefore, a survey was conducted at the interaction field. It is a possible approach that we can get a degree formed mental model. The measurements of mental model were performed according to the procedure described as follows.

II. METHOD

A mental model survey using a questionnaire and a structured interview were executed.

The questionnaire prepared by the measurement of mental model was conducted on the user in a lecture room at the college. This is an analysis system that measures the factors of Japanese mental models and converts them to an index as numerical values.

The following two points (Figure 1) are enumerated as the main points of the idea [8]. Functional model is a kind of a sequential plan in operating machines, such as, "do A, then B, then C", and so on. Structured model is a selective type of plan in doing machines, such as "if B is the D, do that".

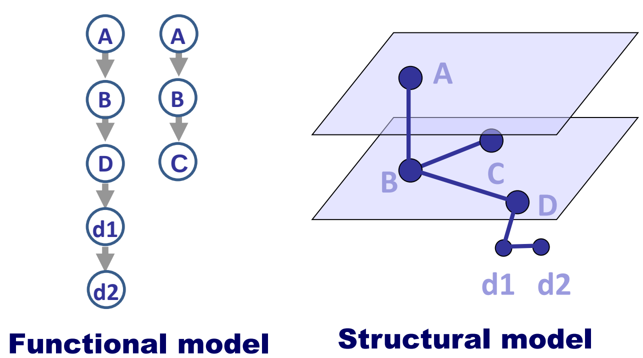


Figure 1. Mental model

C (components) level (Figure 2) is related to Structural model. "Choices" is key for good user interface (terms, visual cue are important).

P (process) level is related to Functional Model. Terms, visual cue are important (special operation such as long press, dual press) (Figure 2).

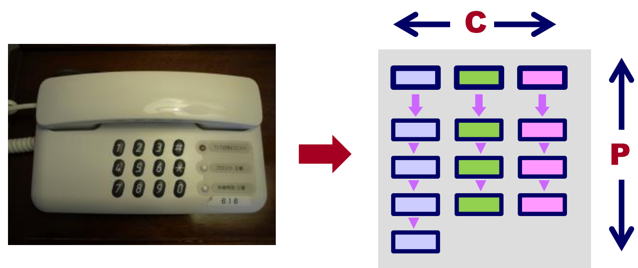


Figure 2. C level and P level

The mental model is operated based on the above-mentioned two points. It evaluates 13 question items in stages.

In order to understand the structure of the mental model, it is necessary to clarify contradictions between elements that prevent the effectiveness of the mental model.

We used the method based on memory of colors and auditory recollection in order to gather accurate information for our research,. The procedure is described as follows.

We performed interviews with 42 people (2 males, 40 females); the participants ranged in age from 20-22 in Japan. We only got to interview 42 people. We asked Nagano Prefectural College to find people for our survey. All participants are pursuing studies in life science as either a major or minor.

A structured interview was conducted with the participants. The questions consisted of product categories, such as faucet, calculator, alarm clock, kitchen timer, etc.

The questionnaire form, pens, and pencils were distributed. First, we chose the user or subject at random. We explained about objects and rules of the interface to each person, and then practiced once with them.

We demonstrated one example interface so as to familiarize the subject with the object and rules of interface sheet; the subjects operated by the themselves (Figures 3-15). After the questionnaire was completed, the forms were collected.

The measurement employed real interface picture. In Figures 9-11, the subject guesses the operation of their mock product. This is followed by a brief statement of the reason for their choices.

In Figures 13-15, the subject follows the same procedures as in Figures 10-12. The subject is asked why they guessed their particular buttons.

In Figures 4 and 5, the subject must guess one specific button from the complete deck. These functions are arranged in a designer's pattern. Figure 6 has only 6 labels facing up.

The procedure is described as follows.

We said, "Please write the procedure shown below

[A] →[+] →[B] →[---]".

Task 1:"Which direction to turn handles to get hot water?"

Task 2:"Show the procedure"

2-1: 2 + 3

2-2: 1000(yen) + tax

 ISBN: 978-1-61208-325-4

2-3: $\sqrt{4} + 2$

Task 3: "Show the procedure"

3-1: 2 + 3

3-2: 2 × 3

Task 4: "How to set the time? Please change 2:17 to 11:45"

Task 5: "What is "reset"?"

Task 6: "How to set the time?"

Task 7: "How to set the time?"

After doing with each subject, we showed them three versions of a mock interface design that we created, which combines television, radio, and alarm clock. Our intention was to analyze the subject's various procedures when asked to perform any given task on our interface designs.

As in Figures 10, 11 and 12, the subject guesses one specific button from those facing down. We said, "Guess the numbers of the one or three mock buttons facing down and state the reason behind your choices. Besides, there are 12 buttons with labels facing down."

The interfaces are designated as task 8 or 11, 9 or 12, and 10 or 13. Subjects numbered the sheet to check starting from first and going to the end.

The experimenters randomly distributed one of two surveys to 42 members of an undergraduate class. 42 of the subjects completed a survey on the "functionality of control buttons on a product".

Designer's model on the interface was set to the procedure on the task 9 and 12; "power" button to "channel" button, "TV/radio" button to "channel" button.

Task 8 or 11; "Please image on the air. Listen to the news program by the use of the radio. How do you operate when listening or doing the radio?"

Task 9 or 12; "Please image on the air. Listen to the news program by the use of the TV. How do you operate when watching or doing the TV show?"

Task 10 or 13; "How do you operate when setting the alarm clock at am 6:30."

The surveys contained a series of graphics representing control buttons. The labeled product group was asked to complete the same survey to determine if product controls were uniformly understood, as well as to provide a comparison with the No-labeled group.

The participants were asked to write down the function of each button next to the graphic.

Figure 3. Faucet

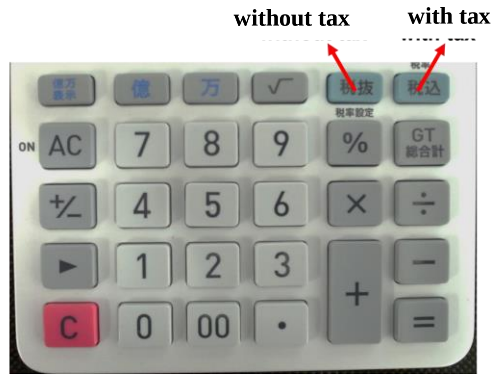


Figure 4. Calculator1

Figure 5. Calculator 2

Figure 6. Alarm clock

Figure 7. Kitchn timer

Figure 8. Alarm clock1 in the hotel (No label)

Figure 9. Alarm clock2 in the hote

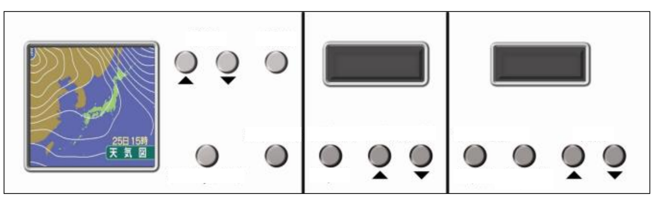


Figure 10. Setting the radio (No label)

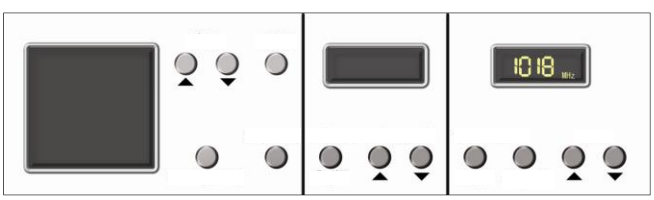


Figure 11. Setting the TV (No label)

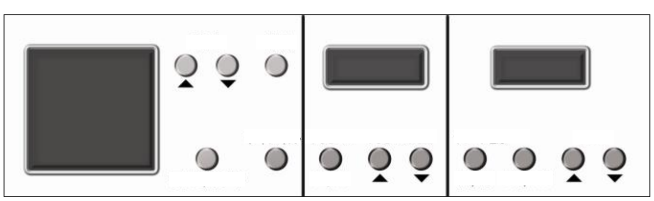


Figure 12. Setting the alarm clock (No label)

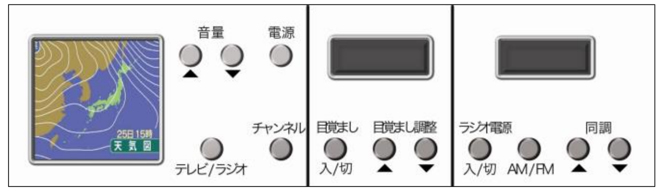


Figure 13. Setting on the radio

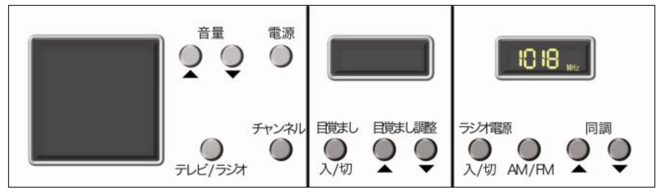


Figure 14. Setting on the TV

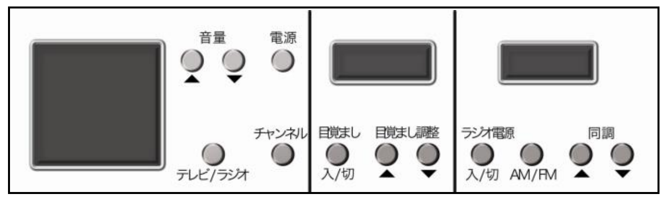


Figure 15. Setting on the alarm clock

III. RESULT

As a result, the formation of a mental model was observed for these interfaces at each task.

In task 1, 100% of the people guessed everything correctly. In task 2_1, 88% of the People guessed correctly. Cues were due to understand the direction to warm water. User's habit caused incorrect mental models to be formed.

In task 2_2, the mental model measures 90%. The other 58% use the correct pattern to guess the correct model in task 2_3. The other 42% did not use the correct pattern to guess the correct model. In task 2, this appears to use practice procedure; "$\sqrt{}$" button to "4" button.

In task 3, almost all the subjects answered inappropriately. The mental model measures 50%. Participants could not practice because of no knowledge about the device.

In task 4, 95% of the people guessed correctly at C level. The other 3% of the people guessed correctly at P level. The mental model measures 49%. 51% did not use the correct pattern to guess the correct model. Participants could not realize how many times they pushed the button; "mode / set" button.

In task 5, the mental model was determined to be 88%. The other 12% did not use the correct pattern to guess the correct model. We almost could found means on how to use it; "reset" button.

In tasks 6 and 7, the measures of No-label and label were determined to be 52% and 56%, respectively. It seems that this result did not depend on labeling for a mental model.

In tasks 8-13, the degree of formed mental model of labeling interface is larger than that of no label interface. It is probable that mental model was formed by the cues with labeling.

These results showed that when trying to guess the buttons rather than using their logic, they used their intuition, or they followed their instinct. Frequently, interface rules escaped their memory by themselves. The rate of correct answer is determined by these factors: the number of button and the alternatives.

Tasks 8-10 and tasks 11-13 show the rate of correct answer cue or alternative of each case in the mock interface, for example, the investigation revealed that 30 out of 42 subjects acted correctly. The rate of information was obtained by changing from the amount of cue. These results show that there is a marked increase in the rate of incorrect answers. The characteristics of such cues are assumed to be the origin of the increase of rate. Considering the previous results, this result was expected.

TABLE I. TASKS 1～7

	Task									
	1	2_1	2_2	2_3	3_1	3_2	4	5	6	7
C level	100	100	95	44	100	100	95	88	29	12
P level	100	76	85	72	0	0	3	/	75	100
	100	88	90	58	50	50	49	88	52	56

TABLE II. TASKS 8～10

	Task				
	8	9	10		
C level	48	40	0		
P level	90	88	–		
	69	64	0	44	Total

TABLE III. TASKS 11～13

	Task				
	11	12	13		
C level	71	60	31		
P level	77	100	15		
	74	80	23	59	Total

IV. DISCUSSION

This research aims to clarify an essential problem of the mental model, and to present a plan for its solution. The measurement approach to determine a solution based on this case was examined. User's mental model is reflected in that result. However, increasing the accuracy of method requires increasing many case studies.

Each of the 42 people selected played with ourselves and on their own. These results provided evidence that people have an illogical jump or mistakes. People might think that truth may exist, but they maintain that if it does, the mental model is incapable of attaining it. Humans may be narrow-minded and refuse to consider certain alternatives only because these alternatives do not meet their prejudgements assumptions about what is and what is not worth pursuing.

In our experiment, from the results in Figures 10-15, one can see that the people adjusted better to the pattern based on habit rather than their pattern. In other words, they did not adjust to a pattern. Moreover, the people save their cognitive resource, and have the characteristics called as "fixed action". When we changed from a pattern to another, with many buttons faced down, this caused a lot of stress for the subjects when having to guess the process in the pattern. The maximum amount of inappropriate cues indicates that the people's thought process is over the allowable maximum of cognition and they get an illogical jump. This should be the optimum amount cue.

This allowed us to further evaluate the functionality of our designs and understand how and why our subjects make certain decisions. By designing a product that is basic and easy to use based on our finding reduces the amount of stress on the mind, and makes it easy for the person to use.

We, however, have to think it over relate a trust, emotion, expectation, situation, human-relationship, etc. Human-computer interaction which should be accompanied by the appearance of mental model cannot be observed in the user's brain. In the future, this may provide a different result. Another case requires a more detailed discussion.

This measurement provided a reasonably good method for such assessments. The method offers many advantages over the conventional method. Applications include on-site testing of paper prototyping.

The mental model approach allowed us a glimpse into the mind of our subjects - that is, we were able to understand their mental model when faced with the decision making which is in fact one of the most important factors of universal design itself. Therefore, it was crucial that we guess the logic behind each person's thought process.

V. CONCLUSION

In conclusion to our research on decision making using the mental model check sheet, our results showed that people have illogical thinking when making a guess based on their mental model. It is likely that people would guess illogically, so in order to support their decision making we should design an interface that is suitable for this age group. In order to be successful with interface design, we must assess user's decision making. For the findings in our research, we proposed a method and gave an answer to this problem based on a concept "thinking or thought". This new proposal suggests "How to design for user" on the interface design of products.

We confirmed that our process had the effective to provide values. This method contribute to the improvement of the usability of interface.

References

[1] R. Mori and T. Yamaoka, “Grasping the Use of Products on Elderly People and Their Mental Models with A Mock Interface Design”, Japanese Society for the Kansei engineering, vol. 8, no.1, pp. 185-195, 2008

[2] J. Nielsen, Usability Engineering, Academic Press. San Diego CA, 1993.

[3] N. A. Stanton and M. S. Young, Guide to Methodology in Ergonomics: Designing for Human Use, CRC Press, 2002

[4] T. Doi, S. Tominaga, T. Yamaoka, and N. Yukiko, “The Elements for Structuring the Mental Model in Operation of User Interfaces,” Japanese Society for the Science of Design, vol. 58, no. 5, pp. 53-62, 2012.

[5] T. Doi, K. Ishihara, and T. Yamaoka, “The Proposal of the Level of Mental Model Building Measurement Scale in User Interface,” Japanese Society for the Science of Design, vol. 60, no. 4, pp. 69-76, 2013.

[6] D. Gentner and A. L. Stevens, Mental Models, Lawrence Erlbaum Associates, Hillsdale New Jersey, 1983.

[7] N. A. Donald, The Design of Everyday Things, Doubleday/Currency. New York, 1988.

[8] J. Preece, Human-Computer Interaction, Reading MA, Addison-Wesley, 1994.

Travel Experience Cards: Capturing User Experiences in Public Transportation

Alma L. Culén, Maja van der Velden and Jo Herstad
Department of Informatics, Group for Design of Information Systems
University of Oslo
Oslo, Norway
Emails: {almira, majava, johe}@ifi.uio.no

***Abstract*— The paper presents a Travel Experiences Card (TEC) set, produced as a tool to understand user experiences in public transportation. The card set allows exploration of diverse user experiences, from preparing travel to arriving at the destination. The set is designed using images taken during participant observation and photographic documentation of public transportation experiences. Two ways or working with the card set are proposed, one based on forced association technique and the other on focus event tool. These allowed for the breadth and the depth of exploration, respectively. We tested the cards in one pilot workshop to fine-tune it as a tool, and then applied it in two workshops in order to evaluate its usefulness. We used three important experiences for users of public transportation: safety, joy, and arriving on time. We found that the method was useful for eliciting rich responses from our participants, and for understanding the ecology of experiences. We also consider the process of designing the card set, and its evaluation, to be of relevance.**

Keywords—service design; interaction design; experience design; customer journeys; experience design cards.

I. INTRODUCTION

Service Design (SD) has gained popularity as an approach to modern business, design research and design education [1]. Like Human Computer Interaction (HCI), SD is a multidisciplinary field. It gained momentum with the introduction of design thinking [2]–[5] some years ago. Many interaction designers have moved into service design field [6], and some HCI designers, including user experience designers, are now involved in the design of services and related research. The evolvement of service design also sparked strong interest in user experience (UX) with services and ways of working with those experiences. In service design, the term *customer experience design* is understood as a holistic concept, which integrates all aspects of the service. The service may be provided by one or more companies, but is considered as one service, as long as the customer experiences the service as one [7]. Of course, every encounter with a product or a service that we use is an experience. *Heterogeneity*, as a characteristic of the service, is a consequence of both the diversity in delivery by a provider and the diversity of people experiencing a service. Customer experience design strives to craft experience that matter and to control a number of factors that shape that experience [8]. The more positive the experiences are, the more loyal the customers will be to the service provider [9]. This has, ultimately, consequences for the company's long-term competitive advantage, as well as on the shaping of values in business other than profit.

Designing good public transportation services is neither a small task nor an inexpensive one. It may be so that "*Being able to profit from innovation is a major incentive to invest in innovation activities regardless of the type of innovation"* [10]. However, service innovation and design, including customer experience design, depend to a much larger degree on collaboration with external parties [11]. Among those parties are universities and design communities aiming at increased benefits for the society as a whole, not just profit. More and more frequently, customers, or users, as we will also call them, are given a stronger voice in the design of services, and service design is often characterized by close work with users and users' involvement in design processes.

In this paper, we continue looking at ways of improving user experiences with public transportation. In [12] and [13], we have considered design for increased visibility of travelling information and ways of gathering and analysing data containing users experiences with the system. In this paper, we take a different approach. We consider using service design and user experience design approaches to describe customer journeys and sets of related experiences.

Our concern is that of methodology: 1) how to engage users in discussions around services and experiences in ways that can promote "good" service and experience design, and 2) how to make use of a "good" tool, such as cards, which are frequently used in HCI [13]. In line with Clathworthy's approach [15] to service innovation, we have designed a tool, a set of travel experience cards, to help us gather relevant information about user experiences while travelling.

The card set addresses all stages of the travel with public transportation, from planning the trip to arrival to destination. The purpose of the tool is to provide data on user experience in initial phases of new service design. The tool was tested in a pilot workshop, then re-adjusted and applied in further two workshops. The set could be used to create customer journeys, address touch points the journey contains, and experiences between the touch points. We have particularly focussed on the latter. The goal of those two workshops was to evaluate the card set as a tool, primarily for its usefulness. The paper reports on the results regarding the evaluation of the tool.

 ISBN: 978-1-61208-325-4

The paper is structured as follows. In Section II, we provide background material on service design, customer journeys, touch points, experience design, and introduce the ecology of experience. In Section III, we present our case, the development of the Travel Experience Card (TEC) set and a way of using it. Section IV provides discussion on the use, and usefulness of TEC. Section V concludes the paper and addresses the future work.

II. Customer Journeys and Ecology of Experience

Service design and customer experience design draw on methods from various fields; several of these methods are familiar to HCI researchers, such as scenarios, role-playing, personas, card sorting, focus groups, and observations. We therefore provide definitions for those methods that HCI researchers may be less familiar with. Service design uses more explorative ways to challenge problem areas, in contrast to HCI design with its more analytic and positivistic approach [16]. In this section we define the main service design concepts we use in the paper.

Defining *services* is challenging, as there exist multiple and competing views. For the purposes of this paper, we view a service as an exchange between a service provider and a user in a particular moment of time and space. Lavrans and Polaine [6] define *service ecology* to be a diagram of all the actors affected by a service and the relationships between them, displayed in a systematic manner. Using service design as an approach to design of a service implies: *"…designing services that are useful, usable and desirable from the user perspective, and efficient, effective and different from the provider perspective. ... service design takes holistic approach in order to get an understanding of the system and the different actors within the system"*, [1, p. 3].

Service design may be considered as "*design for experiences that happen over time, and across different touch points"*, a definition given by Clathworthy [15].

A. Customer Journeys

Customer journey is one of the most effective tools in service design. It is similar to storyboards and use cases in HCI, helping to visualize a service in an organization or company. In [16], Koivisto explains customer journeys as follows: "*Services are processes that happen over time, and this process includes several service moments. When all service moments are connected, the customer journey is formed. The customer journey is formed both by the service provider's explicit action as well as by the customer's choices*".

We consider customer journeys to be formed not only by service moments, but also include all the experiences within and between those moments and user's responses to those experiences.

B. Touch Points

A customer journey is comprised of touch points, the service moments as described by Koivisto [16], or nodes in a visual, graphical representation of a journey. A touch point forms a link between the provider and a customer, and as such, is the origin of customer experiences with service in question. Touch points form one of the three pillars of service design [17, p. 142].

While touch points are a fundamental part of service design and a starting point in re-design of services, we consider the intervals between them to be important for user experience design.

C. Understanding travel experiences and their ecology

An approach to understanding experiences may be that of Nardi and O'Day [18], who use the term 'information ecology' to describe an interrelated system of people, practices, values, and technologies within a particular local environment. This ecology approach, applied to service ecology [6], and the framework for studying user experiences while interacting with technology developed by Forlizzi and Battarbee [19], shaped our theoretical perspective. *"Experiences and emotions are not singular events that unfold without a relationship to other experiences and emotions"*, [19].

Building forth on these understandings, we define *ecology of experiences* as an interrelated, scalable set of experiences along a particular customer journey. In this paper, the context for creating customer journeys is that of travelling with public transportation.

D. Service Design Cards

A good tool for helping to address touch points in the initial stages of service development is a set of service design cards, see [14]. Clathworthy provides six different use contexts for the cards and evaluates the cards related to their intended function. The cards were found to help with team building in cross-functional teams. Further, they were found to be helpful in assisting with the analysis and mapping of existing situations, generating ideas for new solutions or approaches, needs elicitation and facilitation of communication.

III. Development of Travel Experience Card Set

The service design cards described above provided the inspiration for the user TEC set. The modified card set was recently successfully used in another project [20].

Tangible objects, such as cards, and the images depicted on them, are known to facilitate visual thinking and help with finding a common language for communication within groups of people with diverse backgrounds. The common understandings are built through negotiation and discussion of associations and concepts related to images [21], some of good examples of use of cards can be found in [22], [23].

We next describe:

- Design of the TEC set
- How can TEC set be used

A. Design of the TEC set

We started from the perspective of being users of public transportation ourselves. We used participatory observation and photographic documentation [24] to record our own and other travellers' experiences. Armed with cameras, two of the authors went traveling and collected a large amount of

 ISBN: 978-1-61208-325-4

pictures documenting what they considered to be meaningful moments representing public transport traveling experiences of various kinds.

The next step was to map their customer journeys and sort the images of experiences [14] according to touch points on the journeys. However, we found out that it would be more meaningful to consider segments of all customer journeys. All users of public transportation plan their journey in some way. Perhaps the starting touch point for the journey is a smart phone app. The next touch point may be purchasing the ticket on the smart phone, or walking to the tram. Whatever touch points on a particular journey are, they all consist of the subset of the steps shown in Fig. 1. These steps are: planning the trip, making sure one has a valid ticket, arriving to a stop, embarking, traveling (this can be interrupted by, for example an accident, a ticket control, or other forms of disruption), disembarking, perhaps repeating some of the steps if transfer was needed, arriving to a final station and arriving to a final destination. The images for the cards in Fig. 1 were not home-made, rather found on the net, intentionally somewhat different in style than the images we collected. A purple colored stripe was used to further differentiate these cards.

Figure 1. Cards with purple stripe represent segments of a trip

The images collected as representations of touch points and user experiences in public transportation were then sorted into categories corresponding to the segments of customer journeys. There were too many cards in each category. We reduced the number of images by selecting images perceived to be meaningful for everyone, or choosing one image among several that best represented the user experience (the photographers were not part of this process). One or two words were chosen for each image and typed on red background. The first set of TEC cards was thus made, consisting of two different types of cards, those representing segments of the customer journey, and a mixture of touch points and experience cards.

In order to ensure that images convey appropriate experiences and that the text is suitable, we have done quick-and-dirty user testing: we have simply shown the cards and asked two students, who are public transportation users, what they see on cards and if words match what the image conveys. At this stage, we did not want a perfect set of cards, but rather, the one that that was open for modifications and additions. We, for example, chose not to make separate cards for embarking and disembarking, even though one of our testers suggested it. We wanted to see if distinctions in experiences between these two segments were important for users. If they were, separate cards would be designed for the final set.

B. How can TEC Set be Used?

The set of experiences on a customer journey representing a trip using public transportation is very rich. In order to start understanding how to use them and what we can expect to learn from the use, we decided to choose a small set of experiences and focus instead on the use of cards as a tool for working with selected user experiences effectively. Three experiences were proposed as workshop themes: safe travel, arriving in time, and joy while traveling. These were chosen on the basis of the experiences collected during the participatory observation phase. The purpose of using these three experiences was to find ways in which the cards could best be used, in order to gather and understand information that could serve as the basis for designing better travel experiences. We did not study these experiences themselves.

Figure 2. Touch point cards, such as a smart card, and experience cards, such as feeling safe, as they relate to the trip segment Station.

There are two components to the TEC card set: the TEC *cards* and the TEC *tools*. Tools are the ways in which the cards are used in working with users. We have worked with two tools. The first TEC tool is based on the *forced association concept*, and is carried out in relation to every card representing a segment of the journey. This amounts to nine rounds of forced associations, as there are nine cards describing the segments; Fig. 1. Thus, experiences of safety, joy and being on time, each explored in one workshop, were considered across all segments of a journey. A summary of the workshop results can be found in Table 1.

 ISBN: 978-1-61208-325-4

TABLE I. USING TEC TO LEARN ABOUT EXPERIENCES

Stage	Experience		
	Safe	*Joy*	*On time*
Plan the trip	Getting the right information	Coffee Planning software works Having choices Sitting comfortable Access to maps Sheltered Nice view Not crowded Cleanliness	Information about destination Access to maps Buy ticket online/app saves time Having ticket ready
Valid ticket	Getting the right ticket Feeling safe paying for the ticket	No line-ups at ticket machine Working card reader/Feedback from reader Convenient location Hassle-free app Fast service	No line-ups at ticket machine Find the ticket reader Working card reader/Feedback from reader Buy e-ticket while already in buss/train Charged mobile phone
Station	Finding your way Physical safety	Feeling safe No waiting No crowd Good coffee Frequency of departures No special users (no baby carriages, no wheelchairs) Cleanliness No smoking Find platform easily Access to toilet Protected from bad weather	Find platform Map of station/metro Impossible to hear information via speakers Information about destination Information for visually impaired Information about disruptions
Ticket control	Aggressive behavior of ticket controllers	Having a valid ticket No disruption	Disruption Make you miss your stop
Embark /Dis-embark	Validity of the ticket Getting in the right buss/tram etc. Getting off the right stop/station	Driver greets travelers Easy to get in/out No crowd	Too crowded to get in Delay because of children, people with bicycles or stroller, people in wheel chairs
Travel	Fellow travelers Sexual harassment and other violence	Enough space Fast No crowd Information about destination Feelings safe Drivers (?) Charging battery Wifi access	Disruptions
Dis-ruption	The unknown	Speakers that work Waiting area Emergency exit relevant information	Information via speakers Emergency exit
Arrival Stop	Finding way to destination	Mind the gap Finding way to destination Maps No crowd/strollers Recycling of garbage Access to toilet	Map about location Finding way out of the station Valid ticket to get out of station Working card reader Ruter app to check route to destination
Arrival des-tination	Feeling safe, street crime	Being on time Price (of getting to destination)	Time

Figure 3. The cards in use. The cards depicting experiences and touch points are used in conjunction with every segment of the trip, addressing just one kind of experience at the time.

The process of association required about an hour each time, with 3-5 participants; see Fig. 3. We needed to use convergent and divergent conversations, opening up for stories, reflections and memories, but also sense making of these, now collective, experiences.

The second TEC tool we called the *focus event*. A specific, significant event from a person's life, related to the use of public transportation, is placed in the focus of discussion. By significant, we mean an event that is out of the ordinary, either positive or negative. For example, losing a wallet on a city bus, with driver's license and a whole lot of other important documents, would be significant. The cards, both segment and experience/touch points, relevant for a focus event are found and their influence on the event is discussed. The focus event is based on the same experience as the rest of the workshop in which it is used. The process around the focus event takes approximately 5 minutes per event. Again, we were interested in evaluating usefulness of cards for exploring experiences and services related to such events.

IV. DISCUSSION

The first workshop was a pilot workshop where the three authors took time to try different versions of tools, to discuss how it feels to work with touch points and user experiences vs. segments of the journey and user experiences. The protocol that we decided on was to go first through the experience of *safety* in public transport by forced association, followed by a discussion related to a focus event. The subsequent workshop on *joy* as an experience in public transportation was done with two of the authors and three users of public transport, two of them PhD students, and the third a master student in design, use and interaction study program. The last workshop on *arriving on time* included the authors, and two students, one PhD and one master student in the same program.

A small number of participants were involved in the workshops, but they all had a solid background in both user experience design and various methods of working with users, including co-design, participatory design and user-centered design. This is relevant because most of them have worked with similar methods before and could give qualified opinion about the tool, both from the direct experience of using the cards themselves and from their prior experiences of working with users. We felt that five participants with such background were sufficient for giving us the feedback on the TEC set and the tools we chose (the forced association and the focus event) to address TEC's perceived usefulness for design of better user experience when taking public transportation.

A. On the use the cards in workshops

The workshops with our participants started with explanation of the goal of the workshop, the TEC set, and what it is to be used for in the future. We then asked the participants to focus on what gives them, as users of public transportation experience of joy or arriving on time. During the workshop on joy, it became clear that we were missing several experience cards: weather, space, valid ticket, toilet, charging battery, time, and price of ticket. During the workshop on arriving on time, experiences that help users, or are a hindrance to reaching their final destination on time were considered. During this exercise two new experience cards were proposed: *ticket control* and *event*. Ticket control was perceived as both a segment card and an experience card. The event experience card refers to large events, such as sports championships and matches, in which large crowds of people use public transportation. During these events it is often impossible to arrive on time.

By constructing common understanding and meaning giving to the cards, for the entire length of the trip, we found that a number of combinations of experiences have emerged as important. For example, a card with a term 'crowd' was used extensively. It was related to several segments (station, ticket, embarking and disembarking and traveling) to both feeling of lack of safety, lack of joy and danger of being late. One of the participants then mentioned that there is really nothing one can do with this knowledge. This started a whole discussion on the strategies that people use to avoid crowds. At the end of the discussion, all participants agreed that, actually, there are opportunities for making things better by design.

The same conclusion was reached regarding the use of cards to address focus events. We illustrate this with two examples. One participant told a story of a woman who had a very unpleasant experience on the train. She never enjoyed taking public transportation again, and never took trains very early in the morning or late in the night. The card that she held while talking about the experience is 'feeling safe',

 ISBN: 978-1-61208-325-4

'feeling safe', depicting station in dusk, empty, and not giving the feeling of being safe; see Fig. 2.

The second example had to do with embarrassment over being caught without valid ticket and blaming several touch points involving technology that were not working properly at the time. In both cases, cards representing related experiences were found and participants considered frequency of such events, their impact on people's lives, possible design solutions etc.

B. Vocabulary

It was important for the participants to understand the TEC cards. Only then could they really engage in working creatively with them. As it was not possible to have a card representing each individual experience, the terms describing the cards were chosen with care. We found out that some cards needed to be broad enough to allow for several different interpretations. Others, as for example 'ticket' needed further specification: *valid ticket* and *price of ticket* were the requests from our participants. There was also a suggestion to further specify attributes relevant to the validity of the ticket, such as the visibility of information.

C. User's comments

After each workshop, a few minutes were set aside for asking the participants about their experience with the TEC cards. One of the participants (male, 39) said: *They were good to get the conversation going and explore different topics in a quick and easy manner.* Another participant told us: *The images put you kind of into a memory lane. When I look at the station card, I remember my own station and I can feel the experiences. They make me more aware of the things I should think of. I would never come up with as many examples of experiences as we jointly did* (female, 27). Asked whether it was boring to repeat forced association technique, the participants agreed that it was a good experience, connecting the detailed pictures around each segment card into a larger picture which was more relevant: *This was actually a learning experience for me* (female, 26).

V. Conclusion and Future Work

Understanding user experience is important in design of interactive products and services. People's experiences with public transportation, even with a single touch point, such as ticket validation, are very different. Heterogeneity makes working with user experiences challenging.

The TEC set was found to respond to this challenge adequately. Heterogeneity remained visible, yet a common understanding of experience one focused on during the workshop emerged.

The size and the feel of cards were found to be satisfactory. Part of their appeal was attributed to images. The images were taken out in the field, but were generic enough to easily evoke memories of many diverse experiences. The other part of the appeal was tangibility of the cards. They served as tangible pointers to experiences, evoking memories and facilitating conversation about experiences. They enabled rich communication, in depth when working with focus events, and in breadth when working with forced associations across all segments of a customer journey. Our focus was not on re-designing services at this time, yet, many ideas and thoughts that emerged on during the workshops would be worth pursuing further.

While these conclusions are in line with previously published work and thus not revolutionary, we hope that we have explained the process of creating the tool and its evaluation (by users of public transportation, with good understanding of design processes) in such a manner that it is inspirational. This method of creating a set of cards for studying user experiences is rather fast and fun. The set can be used to understand a range of experiences in a given context of use, both in breadth and in depth, identifying clear design and innovation opportunities.

We have tested a small number of TEC tools (forced association and focus event) and we will further explore other possible tools. Our future work also includes scalability, and different application areas. In terms of scaling, we are interested in two types. The most obvious one is to use the experience cards in much a larger setting, with large number of users. However, we are also interested in understanding the scaling of experiences themselves, how they become larger and larger until the small ones are forgotten. Finally, we want to bring this method into other application areas, such as service innovation and customer experiences in the library, or with young patients in transition from children's hospitals to adult ones.

References

[1] J.-V. Nisula, "Searching for definitions for service design - what do we mean with service design?", ServDes 2012, Co-Creating Services, 2012, pp. 171–175.

[2] R. L. Martin, Design of business: Why design thinking is the next competitive advantage. Harvard Business Press, 2009.

[3] T. Brown, "Design thinking," Harward Bus. Rev., vol. 86(6), 2008, pp. 84–95.

[4] T. Brown, Change by design: how design thinking can transform organizations and inspire innovation. HarperCollins Publishers, 2009.

[5] A. L. Culén and M. Kriger, "HCI in IT-facilitated Business Innovation: a Design Thinking Perspective," in Proceedings of HCII 2014, in press.

[6] A. Polaine, L. Løvlie, and B. Reason, Service design: from insight to implementation. Rosenfeld Media, 2013.

[7] L. G. Zomerdijk and C. A. Voss, "Service Design for Experience-Centric Services," Journal of Service Research, vol. 13 (1), 2009, pp. 67–82.

[8] J. Teixeira, L. Patrício, N. J. Nunes, L. Nóbrega, R. P. Fisk, and L. Constantine, "Customer experience modeling: from customer experience to service design," Journal of Service Management, vol. 23, no. 3, 2012, pp. 362–376.

[9] J. J. Garrett, "Customer Loyalty and the Elements of User Experience," Design Management Review, vol. 17, no. 1, 2006, pp. 35–39.

[10] P. Hurmelinna-Laukkanen and P. Ritala, "Protection for profiting from collaborative service innovation," Journal of Service Management, vol. 21, no. 1, 2010, pp. 6–24.

[11] H. W. Chesbrough, "Why Companies Should Have Open Business Models," MIT Sloan Management Review, vol. 48, no. 2, 2007, pp. 22–28.

[12] M. van der Velden, A. L. Culén, J. Herstad, and A. Atif, “Networked Visibility: The case of smart card ticket information,” ACHI 2013, The Sixth International Conference on Advances in Computer-Human Interactions, 2013, pp. 228–233.

[13] M. van der Velden and A. L. Culen, “Information Visibility in Public Transportation Smart Card Ticket Systems,” International Journal On Advances in Networks and Services, vol. 6, no. 3 and 4, 2013, pp. 188–197.

[14] G. Rugg and P. McGeorge, “The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts,” Expert Systems, vol. 22, no. 3, 2005, pp. 94–107.

[15] S. Clatworthy, “Service innovation through touch-points: development of an innovation toolkit for the first stages of new service development,” Int. J. Des., vol. 5, no. 2, 2011, pp. 15–28.

[16] S. Holmlid, “Interaction design and service design: expanding a comparison of design disciplines”, Nordes, 2007, pp. 1–8.

[17] S. Miettinen and M. Koivisto, Designing Services with Innovative Methods. Univ. of Art and Design Helsinki, 2009.

[18] B. A. Nardi and V. O’Day, Information ecologies: using technology with heart. MIT Press, 2000.

[19] J. Forlizzi and K. Battarbee, “Understanding experience in interactive systems,” in Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques, 2004, pp. 261–268.

[20] A. L. Culén and A. Gasparini “Find a Book! Unpacking Customer Journeys at Academic Library”, ACHI 2014, The Seventh International Conference on Advances in Computer-Human Interactions, 2014, ThinkMind Digital Library, www.thinkmind.org.

[21] E. Blevis, E. Churchill, W. Odom, J. Pierce, D. Roedl, and R. Wakkary, “Visual Thinking & Digital Imagery,” CHI ’12, 2012, pp. 2715–2718.

[22] C. M. Barnum and L. A. Palmer, “More Than a Feeling: Understanding the Desirability Factor in User Experience”. CHI ’10, 2010, pp. 4703–4716.

[23] M. Sunnari, L. Arhippainen, M. Pakanen, and S. Hickey, “Studying User Experiences of Autostereoscopic 3D Menu on Touch Screen Mobile Device”. Proceedings of the 24th Australian Computer-Human Interaction Conference, 2012, pp. 558–561.

[24] M. Crang and I. Cook, Doing ethnographies. London: SAGE, 2007.

Coupling Artificial Neural Networks and Genetic Algorithms in Redesigning Existing Cities for Flood Resistance

Gene Oliver Cruz[1], Florencio Ballesteros, Jr.[1], Ariel Blanco[2]
[1]Environmental Engineering Graduate Program
[2]Department of Geodetic Engineering
University of the Philippines
Quezon City, Philippines
gene_cruz@yahoo.com, {balleste, acblanco.updge}@gmail.com

***Abstract* – Cities are becoming more vulnerable to natural hazards due to increasing concentration of urban population and resources as well as to changing weather patterns caused by climate change. Adding to the aggravation of urban vulnerability is the socio-economic conditions of its population. More often, the poor are those who are severely affected and have no economic means to recover. Eventually, the local government takes the responsibility of providing services to restore and rehabilitate affected communities. This impacts the cities' economic base by reducing their ability to grow and raise revenue. In order to minimize economic losses caused by a disaster, it is important to assess the communities' vulnerabilities and plan ahead before a disaster strikes. This paper explores the use of neural networks and genetic algorithms as support tools for an integrated urban development and disaster risk reduction planning and decision making.**

Keywords-artificial neural networks; genetic algorithms; land use; disaster preparedness.

I. INTRODUCTION

Disaster risk is a key threat to dense urban spaces as evidenced by events in the past [1]. Experiences of the United States of America from the devastation of Hurricane Katrina in 2005 only revealed that even in highly developed countries, buildings and infrastructure, flood risk preparations and well-resourced emergency services can be overwhelmed. Such concern is not an indication of climate change yet but rather a proof of vulnerability of urban areas and settlements to extreme weather events [2]. Disaster risk arises when hazards interact with physical, social, economic and environmental vulnerabilities. Disasters, often, hold back urban development and impact the economic base of urban areas by reducing their ability to grow and maintain carrying capacity. Urban areas, particularly in developing countries, are faced with the challenges brought about by climate change. It is foreseen that a significant amount of residential developments will be required which would shape future spatial patterns of these urban areas. It is, therefore, important that the direction of spatial pattern and climate change adaptation of urban infrastructures are planned [3].

Lately, there has been a gradual awakening to the fact that disaster risk reduction and climate change adaptation are not separate specialized fields of activity alongside poverty reduction and sustainable development efforts. The growing concerns on the impacts of climate change bring about the importance of integrating land use planning, disaster management and climate change to address disaster risk issues holistically [4]. One of the things that link them is their common need to understand livelihoods and obstacles to people gaining access to what is needed to stabilize their livelihoods and make them hazard resistant [5]. Understanding spatial and temporal processes of urban development and its corresponding social-environmental consequences deserve serious study due to its direct and imminent impacts on human beings [6]. This study explores an integrated urban land use and disaster risk reduction approach proposing the use of evolutionary computing algorithm for a holistic planning and decision making process.

The paper is presented as follows: Section II discusses the rationale behind the proposed research. Section III presents the temporal and geographic scope. Section IV outlines the research methodology which is divided into four major activities. Last, since the research is still in progress, this section presents initial results and challenges faced in doing the research.

II. RELEVANCE

In the past, land use management and disaster risk management were treated as almost, if not completely, two separate disciplines. The gradual recognition of their relationship to the city's economic growth and development prompts interest to integrate the two fields. Current integrated approach, if there is any, is usually manual and subject to the whims of the planners and decision makers requiring subjective consideration of various characteristics of the city.

Numerous studies exist on computer supported land use simulation [6][7][8][9][10] but approaches that specifically consider disaster management aspects are rare. The research proposes an automated integrated analysis of land use and disaster risk through processing of geospatial and statistical data to better understand and assess the city's current and future state which aids in effective planning and decision making. The proposed approach views disaster mitigation as an integral part of land use management that does not only addresses the current requirements of the city but also the effects of such decisions on shaping future land use patterns and communities' vulnerability.

III. SCOPE

Metro Manila is the primary core of Philippine politics and commerce. It is a result of integrating seventeen previously distinct territories which includes Marikina City located at the eastern border. Marikina City is consist of sixteen communities or *barangays* and is regarded as a highly urbanized first class city. It has high exposure though to recurring floods due to its proximity to Pasig and Marikina rivers. In 2009, the city was inundated by torrential rains brought about by Typhoon Ketsana damaging properties and claiming lives. Since then, the city has been active in disaster mitigation and preparedness.

Marikina City is selected as a case study for assessing the effectiveness of the proposed integrated approach. Using historical empirical data consisting of the site's physical and socio-economic attributes for the past twenty years with a ten year interval, future land use pattern is simulated at an interval of ten years considering the boundary effects of surrounding cities and municipalities.

Vulnerability indices of *barangays* are also measured using same data sets with the assumption that vulnerability degrades as urban growth is realized. While simulating future land developments, the study also examines the effects of such development to the vulnerability of the *barangays* to flooding over time. The study then proposes a physical arrangement of *barangays* resulting to lesser vulnerability. The results from this simulation can be considered during planning for both land development and disaster mitigation.

IV. METHODOLOGY

The study is divided into four parts: 1) assessing the vulnerability over time, 2) developing a land use model and simulating future land use patterns, 3) optimizing the resilience of predicted land use against flooding, and 4) conducting sensitivity analysis and determining effects of controlling attributes on future resiliencies. Figure 1 illustrates the framework of the research design.

Artificial Neural Networks (ANN) [7] and Genetic Algorithms (GA) [12] are used in this study due to their capability of capturing the complexity, non-linearity, heterogeneity, and at times, chaotic and unpredictable behaviour of both urban growth and community resilience. These can accommodate and process volumes of data and identify which of these data are significant. Initially, the study focuses on loosely linking ANN and GA. Each method is independently performed on individual platforms with the intention of integrating them seamlessly in a form a program simulation.

A. Land Use Simulation Model

This subsection undertakes the development of a land use simulation model that defines the probability of land use transition. Historical empirical data are used to determine the relationship between site attributes and urban development. To achieve this, the geographic map of Marikina City is initially broken down into acceptable grid or cell sizes (e.g., 50 meter by 50 meter grid). Future state or development level of each cell is predicted by considering both its current site attributes and development state (residential, commercial, industrial, etc.) and of its neighboring cells. Collectively, the cells generate the predicted land use pattern of the city.

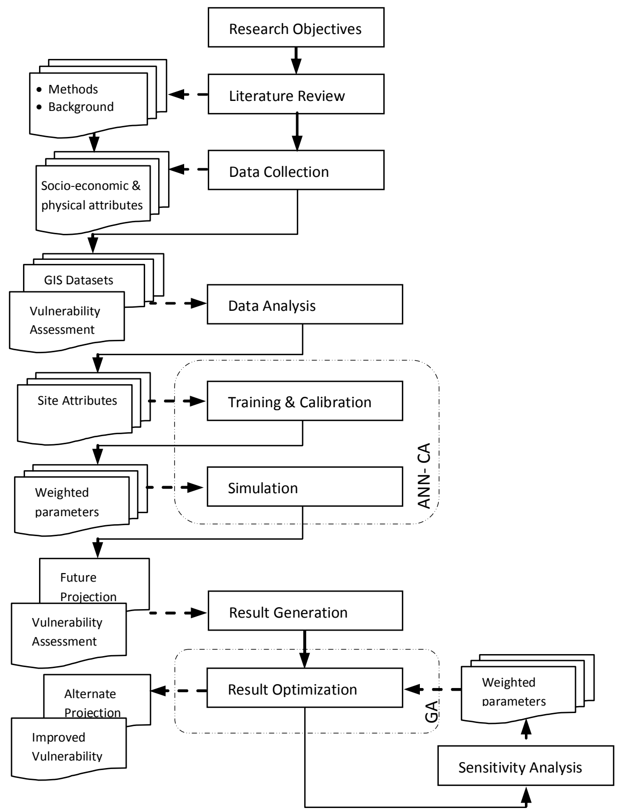


Figure 1. Research Methodology

Geographic Information System (GIS) data layers of site attributes (S'_n) influencing land use change such as socio-economic data (e.g., population growth, population density, etc.) and physical characteristics (e.g., slope, proximity to infrastructures, etc.) from three epochs or periods (1990, 2000 and 2010) of Marikina and its neighboring areas are collected and processed for purposes of data clean-up, conversion and extraction of data layers. Physical attributes related to geographic proximity and accessibility are also extracted using GIS software (e.g., ArcGIS) [10] common analysis functions such as buffer and overlay. Socio-economic statistics are also entered into the software as a data layer. There are two ways in handling the historical site attribute for training. One approach is to assume that development proceeds based on historical trends. The other approach is to modify the data following a set of planning criteria [7].

ANN is employed in determining parameter values for the transitions rules of the proposed land use model. There are two parts of developing the model: training by ANN and simulation by Cellular Automata (CA) [6][7], as illustrated in Figure 2. The neural network is composed of three layers: an input layer, a hidden layer and an output layer. The input layer has *n* neurons representing *n* site attributes obtained from GIS. Similarly, the hidden layer has *n* neurons. The output layer has only one neuron indicating the development probability. At each iteration stage, the neural network determines the development probability based on the input site attribute. The algorithm derived from ANN is used for CA simulation.

In order to achieve the best generalization, the data sets are split into three parts [8]:

- The training set which is used to train the neural network, and its error is minimized during training;
- The validation set which is used to determine the performance of a neural network on patterns that are not trained during learning; and,
- The test set which is meant for checking the overall performance of a neural network.

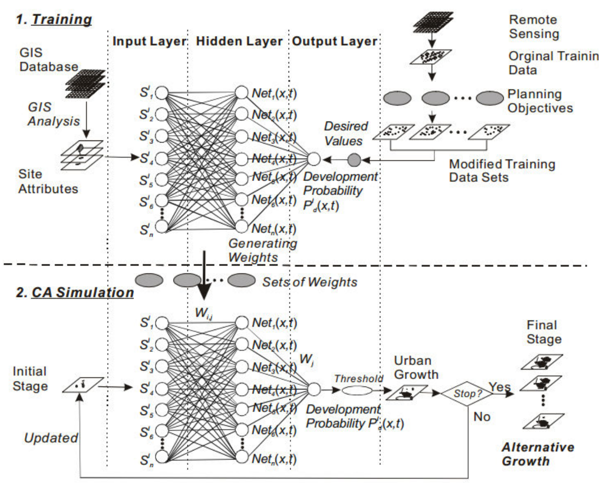


Figure 2. Simulation of urban form using ANN based CA model [7]

Historical data of 1990 and 2000 are used as training data to calibrate the neural networks. Training or learning is then stopped when the validation set error reaches its minimum. After concluding the learning phase, the net should be finally checked with the third data set – the test set, i.e., actual land use of 2010. CA simulations are carried out for the three periods so that land use transitions can be compared and evaluated against the actual development of 2010.

ANN is employed to estimate the development probability at each iteration stage of the CA simulation. The existing variables of a cell and its neighboring cells are the input values and the neural network determines the development probability at the output layer at each iteration stage. Before keying the values into the neural network, variables are scaled to treat them as equally important inputs and to make them compatible with a sigmoid activation function [7]. The parameters or weights are significant in defining the final signals. Back propagation algorithm is adopted as an iterative learning procedure to minimize the error function. In this process, the parameters undergo continuous adjustments by comparing the calculated outputs against desired outputs [9][10][11].

A random seed number, an arbitrary constant determined through trial and error, is determined and used to initialize the weights. The errors, computed as the difference between calculated and desired output neuron, are propagated backwards to refine the weights. The adjustment of the weights is iterated or repeated until such time that the errors are within the acceptable thresholds. This completes the training process of the neural network making it ready for simulation.

Another consideration of this study is to validate also the results against two major analytical characteristics of spatial analysis: spatial autocorrelation and spatial heterogeneity [11]. Once training is completed and appropriate weights are obtained, the data are entered and processed into the neural network based CA simulation model. The model is implemented in a GIS platform by integrating neural network and GIS. A graphical result of the model is presented in GIS format.

B. *Vulnerability Assessment of Current and Future Land Use Change*

Spatial planning with the aid of GIS has also been regarded as becoming one of the most important tools in disaster risk reduction. Spatial data, together with non-spatial ones, can be analyzed in obtaining information on geo hazards and hazard prone areas which can be used as planning and decision making tools.

Using a "weighted average" technique, vulnerability assessment of Marikina's *barangays* is performed to provide insights on the historical trend of their vulnerability. Socio-economic variables are processed into weighted variables using data analysis software (e.g., MATLAB) and used in establishing vulnerability index over periods of time. The results are then compared and validated against other risk assessment models. Using the same method, vulnerability of the predicted land use pattern is measured. The predicted pattern is also superimposed into a geo hazard map to determine which settlements would likely be affected. Its vulnerability is assessed using parameters such as loss of life, property, livelihood and infrastructure.

C. *Alternative Future Land Use Change Scenario*

Using GA, a proposed urban arrangement having the lowest possible vulnerability is generated using a cost fitness function as a chromosome. Desired result is achieved through crossover, selection and mutation of its best attributes and producing a desired attribute. The method is similar to Charles Darwin's evolution of species by survival of the fittest whereby a new individual evolves by the crossover of genetic information of two parents.

The measured vulnerability is compared against current vulnerabilities to determine whether there is significant improvement on the communities' resilience.

D. Sensitivity Analysis

Since proposed arrangements are machine generated results, there is a need to validate and assess their practicality following land use planning principles. A sensitivity analysis is performed by controlling or altering attributes as a representation of current city planning or remediation by entering the modifications into the simulation model. The results provide insights to planners and decision makers on the relationship of attributes and the effectiveness of existing measures to reducing communities' vulnerabilities.

V. Discussion and Conclusion

The study is currently on-going. Historical data have been collected from the cluster of cities and municipalities which includes Marikina City, Quezon City, San Mateo, Antipolo City, Cainta and Pasig City. As expected, obtaining complete, up to date and accurate data poses a challenge. This is primarily due to the inability of the local government to retain historical information, particularly in the 1990s where computers were not yet widely used to store data. Most documents were kept in hard copies and retained for a certain period of time, for example, five years after which these are disposed. As a result, additional work is carried out in tracing back historical land use pattern through differentiating available data sets and focusing only in areas where changes occurred.

The number of parameters or attributes to be considered in land use planning and disaster risk management is abundant. Some of these attributes take first precedence over the others and some of these attributes may not be significant at all. For a human to process all available information, it would take much time and effort and by the time results are derived new sets of change have occurred. The research intends to verify the usability of identified problem solving methods in obtaining valuable information for planning and decision making in an integrated holistic approach.

The research also recognizes that the computational requirements may be extremely resource consuming. As such, the research intends to further investigate in the future the use of other computational methods to compare which are more efficient.

References

[1] S. Hochrainer and R. Mechler, "Natural disaster risk in Asian megacities: A case for risk pooling?", Cities, vol. 1, issue 28, February 2011, pp. 53-61.

[2] S. Huq, S. Kovats, H. Reid, and D. Satterthwaite, "Editorial: Reducing risks to cities from disasters and climate change", Environment and Urbanization, vol. 19, no. 1, April 2007, pp. 3-15, http://eau.sagepub.com/content/19/1/3 [retrieved: February 2014].

[3] K. Otto-Zimmerman (ed), 2011, Resilient Cities: Cities and Adaptation to Climate Change Proceeding of the Global Forum 2010, Local Sustainability 1, SpringerScience+Business Media, Dordrecht

[4] B. Bajracharya, I. Childs, and P. Hastings, "Climate change adaptation through land use planning and disaster management: Local government perspectives from Queensland", Refereed paper presented at 17th Pacific Rim Real Estate Society Conference Climate change and property: Its impact now and later, 16-19 January 2011, Gold Coast, pp. 1-16.

[5] B. Wisner, "Are We There Yet? Reflections on Integrated Disaster Risk Management after Ten Years", Journal of Integrated Disaster Risk Management, vol. 1, no. 1, 2011, pp. 1-14, Published online: 30/04/2011.

[6] Y. Liu, Modelling Urban Development with Geographical Information Systems and Cellular Automata, CRC Press / Taylor Francis Group, Boca Raton, Fl.

[7] A. G. Yeh and X. Li, "Urban Simulation Using Neural Networks and Cellular Automata for Land Use Planning", Advances in Spatial Data Handling: 10th International Symposium on Spatial Data Handling, 9-12 July 2002, pp. 451 -464.

[8] C. M. Almeida, J. M. Gleriani, E. F. Catehon, and B. S. Spares-Filho, "Using neural networks and cellular automata for modeling intra-urban land use dynamics", International Journal of Geographical Information Science, vol. 22, issue 9, pp. 943-963.

[9] C. M. Almeida and J. M. Gleriani, "Cellular Automata and Neural Networks as a Modelling Framework for the Simulation of Land Use Change", Anais XII Simposio Brasileiro de Sensoriamenti Remoto, Goiania, Brasil, 16-21 April 2005, INPE, pp. 3697-3705.

[10] O. Okwuashi et al., "GIS Cellular Automata Using Artificial Neural Network for Land Use Change Simulation of Lagos, Nigeria", Journal of Geography and Geology, vol. 4, no. 2, May 2012, pp. 94-101, http://dx.doi.org/10.5539/jgg.v4n2p94 [retrieved: February 2014].

[11] D. Triantakonstantis and G. Mountrakis, "Urban Growth Prediction: A Review of Computational Models and Human Perceptions", Journal of Geographic Information System, vol. 4, no. 6, December 2012, pp. 555-587, doi: 10.4236/jgis.2012.46060 [retrieved: February 2014]

[12] V. P. Bongolan, F. Ballesteros, K. A. Baritua, and M. J. Santos, "Prioritizing the Components of Vulnerability: A Genetic Algorithm Minimization of Flood Risk", EGU General Assembly, Vienna, Austria 2013, 7-12 April 2013.

Simplified Customer Segmentation Applied to an Outbound Contact Center Dialer

Telmo Ricardo Lopes Alberto, Pedro Mendes da Silva
Research and Development
Altitude Software
Oeiras, Portugal
{Telmo.Alberto, Pedro.Silva}@altitude.com

Abstract— **Contact centers are a critical link between companies and their customers. In this context, outbound dialing represents a major business area for many companies, as they often need to reach a very large number of customers by phone or other media. In this scenario, it becomes a necessity to prioritize which customers should be contacted first, probably according to their expected business value. In this paper, we will keep our focus on customer segmentation for an outbound contact center dialer. This implies targeting different sets of customers (customer segments), with distinct priorities and staff for each segment, while taking into account all the business objectives. Our proposal is based on extending our existing outbound contact center management system to provide support for simplified customer segmentation in the context of outbound dialing. This system provides an interactive interface for handling typical contact center business requirements. Our main focus is enhancing the system's user experience, so that it allows the user to manage customer segments and dialing effectively, by using schedules, key performance indicators, multidimensional statistics, business segment prioritization, customer contact prioritization, and staffing management. Ultimately, we have shown how it was possible to enhance an outbound dialer with customer segmentation concepts, focusing on schedules, KPIs, multi-dimensional statistics, business segment and customer contact prioritization by business value, and staffing. Users of the new module find that their staff productivity and responsiveness to events regarding contact list quality has improved dramatically.**

Keywords-contact center; customer segmentation; dialer; near real-time business intelligence; business applications.

I. INTRODUCTION

Contact centers represent one of the main interfaces between companies and their customers [1]. Many contact centers need a large specialized staff to handle all the communications with the customers. They are currently not limited to phone calls as in the past, and have extended their reach to other media, such as email, chat and social networks [2][3].

In contact centers, we usually have two types of interactions, inbound (originated from costumers) or outbound (initiated from inside the contact center) [4]. For example, an outbound contact center can be focused on dialing calls to a list of customers that have a debt to pay. The calls must then be delivered to contact center operators (also known as "agents").

The work in outbound contact centers can be divided in several Outbound Campaigns, which take lists of customers and a list of business rules and try to deliver interactions between the most adequate agents and the selected customers, according to specific business rules.

Contact center management is a complex and demanding endeavor, with the need to connect business objectives, people and processes. Several business applications have been developed to handle this complexity and allow for optimized contact center management.

Customer segmentation is mainly the action of dividing a customer base into smaller sets of individuals, similar in specific ways relevant to marketing. These segments, allow companies to allocate the most appropriate marketing resources to specific groups of customers, in order to maximize the financial value generated [18]. Examples of implementation of the concept can be found in [8][9].

In the past years, customer segmentation has become a very important business requirement regarding obtaining new customers and increasing customer retention. Following this trend, the need to somehow integrate this concept in customer relationship management (CRM) systems, in general [25], and in contact centers management systems, in particular, arose. The underlying goal is nearly always profits maximization.

In this paper, we will focus on customer segmentation in the specific context of outbound contact center campaigns. Section 2 presents some details about our existing contact center management system. Section 3 details the customer segmentation concept. Section 4 describes how outbound campaign customer segmentation was integrated into our system, with the introduction of the Strategy Manager module. Section 5 presents some conclusions and future work.

II. CONTACT CENTER MANAGEMENT SYSTEM

Our previous research has been focused on developing a Contact Center Management System (Altitude uCI). A general simplified overview of the current architecture can be found in Fig. 1.

Our Contact Center Management System is responsible for handling all interactions and events between customers and contact center staff. The dialer is responsible for delivering interactions to agents, without them having to perform manual dials. It offers support for predictive dialing [4], always respecting legislation (by avoiding silent calls) and business objectives.

All the contact center data (system tables, customer data) is stored in an external relational database management system (RDBMS). Working on top of the external database, we have our own application specific built-in Online Analytical Processing (OLAP) engine [5]. This is needed for real-time monitoring, using OLAP dimensions and cubes, extra details about this can be found in [19].

Our system includes client applications, mainly web and desktop applications for the agents (the contact center operators), and a specific web application for contact center supervisors and managers, designed to monitor all contact center performance in real time and react according to it. It also allows the user to perform all the system configuration tasks, as well as requesting and visualizing reports.

The link between the contact center and the customers is provided by the Public switched telephone network (PSTN) and by the Internet. PSTN provides the link for calls and faxes from/to customers, and the Internet provides interactions through social networks, chat, email, and VOIP (voice over IP).

Our system enables contact center stakeholders to manage the contact center according to business goals, using business intelligence technology and concepts (BI). BI is a collection of decision support technologies for the enterprise aimed at enabling knowledge workers to make better and faster decisions [5].

A Key Performance Indicator (KPI) is a measure or metric that evaluates performance with respect to some objective. It is routinely used by organizations to measure both success and quality in fulfilling strategic goals [17].

The system allows business users to track near real-time business KPI, using visual dashboards. Also, often those KPI are associated to specific goals, and trigger special events or actions, if the goal is either reached or failed. Some industry standard KPIs are built into the system (for example: Conversion Rate [6]).

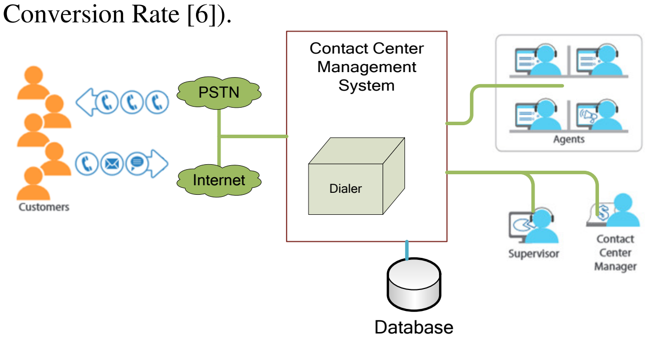


Figure 1. General overview of our contact center management system

III. CUSTOMER SEGMENTATION

The concept of customer segmentation is a key marketing notion. The main goal of separating customers into different segments inside the same campaign is to be able to differentiate marketing activities towards these segments [7].

Extensive research has been done on customer segmentation (both in general and for specific fields), but, ultimately, it is up to the business manager to decide which parameters should be used to perform distinctions between customers and to decide at which segment each customer belongs. Depending on the goal, some simple criteria like gender, age group, geographical location, income, and previous sales are good indicators to be used when segmenting.

For example, “expected profit greater than X” might define a segment. As most companies want to maximize profits (or some other quantity, for example, sales), marketers quickly realized that a segmentation should ensure that better customers are separated from the remaining customers [7]. The concept of customer lifetime value and fidelity is also mostly used in segmentation [8][9].

Regarding the definition of which segments to use, and the attribution of customers between each of them, there are several techniques available. These techniques include K-means (more details for this method can be found in [21]), Two Step, and Kohonen networks [20]. Statistical mixture models, like Latent Class Analysis (LCA), can also be used (a recent example of this method usage can be found in [22]).

Customer segmentation techniques can be generically used for marketing purposes. Some consumer markets customer segmentation types include [20]:

- Value based.
- Behavioral.
- Propensity based.
- Loyalty based.
- Socio-demographic and life-stage.
- Needs/attitudinal.

Industry-specific customer segmentation methodologies have been developed, which include data specific for each field. For example, for the banking industry, there are specific techniques for segmenting credit card holders and specific techniques for segmenting retail banking customers. This also applies to other industries, such as the retail industry, telecommunications industry, and the electrical power industry [20] [10].

In recent years, with the arrival of social media channels, the issue of customer segmentation in this new context has captured research attention. Issues include how to integrate and correlate all the new social data with data from other sources, for example, with market research data and Customer Relationship Management (CRM) data, and also how to match social media identities with all that data already collected in the corporate databases [23][24]. In the

 ISBN: 978-1-61208-325-4

same field, we have seen an interesting proposal for automatic customer segmentation applied to social CRM systems [25].

A correct combination of customer segmentation with the right marketing options for each segment will lead to maximized profits. In the context of contact centers, the main goal is to optimize business results by choosing the right contacts to dial at the right time, maximizing the success/cost ratio of outbound campaigns.

IV. STRATEGY MANAGER

"Strategy Manager" is a component that tries to add simplified customer segmentation for outbound contact center dialers to our system. Visually, it is mainly composed of a visual dashboard (Strategy Center), a calendar (Strategy calendar), and a monitoring area. The user can control the dialer behavior by specifying which segments should be contacted first and which agents should handle the interactions. Near-real time information about the progress in each segment is available, so that the user can act upon it.

Before the integration of this concept in our system there was no notion of business segment associated with a customer list. The only available option was to specify a Structured Query Language (SQL) filter applied directly to the database to select which customers are contacted while the filter is in place. Thus, the user could not monitor progress on each customer segment individually nor use a different staffing for each segment. Priorities between segments were also impossible to achieve without manual intervention.

A. Strategy Center

A matrix-like user interface control was selected after several design and usability test iterations. It was named "Strategy Center", and is shown in Fig. 2. It is a dashboard, showing near real time important business information, and providing access to the business segmentation configuration operations. It is now an important part of our contact center management tool. The matrix is mainly composed of three sections:

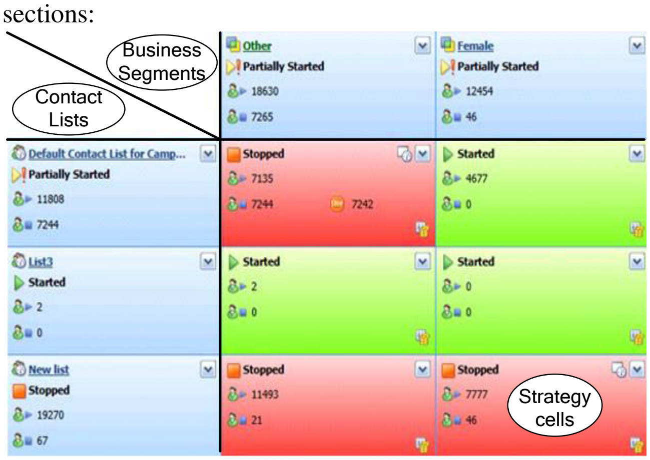


Figure 2. Screenshot of the Strategy Center interface .

- **Contact Lists**: these are the matrix rows. They usually represent different lists of customers to be contacted. Normally, different contact lists represent different input files from customer list data sources.
- **Business Segments**: these are the matrix columns. These contain a filter to specify which customers from the existing contact lists should be contacted.
- **Strategy Cells**: these are the matrix cells, the intersections between business segments and contact lists. They represent the set of customers from a given contact list that match a business segment filter.

With the help of several visual indicators, the user can, at each moment, in each matrix cell, monitor:

- How many customers were already contacted.
- How many customers were not yet contacted.
- If a matrix cell is being targeted (Started or Stopped).
- What is the status of all of the conditions that are currently being used to control the automatic running status of each cell (schedules and KPIs).

More globally, in the matrix top and left cells, the user can also monitor how customers are distributed across the contact lists and business segments. Finally, the user can also monitor the dialing business performance and change the dialing strategy.

B. Business Segments

The main configuration parameter for a business segment is the segment filter. The filter is defined in the user interface, using logical operators, combining customer data fields which can range from demographic data (age, gender, income, etc.) to specific business fields (previous sales, customer category [11] , etc.). It is up to the user to define thoroughly which segments will fit best their business goals, and which customer attributes should be used in the filter to define their positioning.

C. Dialer Management using Business-oriented Rules

The user can fine-tune the dialer behavior by selecting different policies for each strategy cell.

For each cell, there are two main operation modes, the manual and automatic modes:

1) Manual mode: the dialing for a given cell is totally controlled by the user. The user manually selects the "Start" and "Stop" operations.

2) Automatic mode: the dialing for a given cell is automatically controlled by one or a combination of conditions configured by the user, according to specific business strategies:

a) Schedule: the period of time during which the customers from one cell should be contacted.

 ISBN: 978-1-61208-325-4

b) KPI: the dialing for a given cell is controlled by the value and objective of a user-defined KPI.

c) Maximum customers contacted: the dialing for a given cell is controlled by the value and objective of this built-in KPI.

D. Monitoring

It is possible to monitor the dialing process in real time, having individual performance data in every cell, and aggregated statistics for the contact lists and business segments. At each moment, we can tell how many customers were contacted (successfully or not) and how many are left. The monitoring operations are supported by a multidimensional data model [5].

KPIs also enable the user to examine the performance of each cell at any time and whether the business objectives are being met, and dynamically respond accordingly.

Additionally, the user can access more real-time and historical indicators, with options to navigate to other monitoring screens containing tables or charts. Historical-data reports are also available.

E. Staffing

It is often necessary to assign specific agents to specific business segments. For example, a debt collection contact center manager can decide to assign contacts with a larger debt to a set of highly skilled agents. Another example would be to segment customers by education level and let customers with a college degree be handled by agents with a college degree [12].

Strategy Center provides a mechanism for staffing, which includes assigning skills to agents and assigning skills to business segments. The system then matches the required agents to the assigned business segments.

F. Business segment prioritization based on business value

Different business segments can provide different expected business values. For example, consider an outbound contact center where a customer list is sorted in descending order by Customer Lifetime Value (LTV) and then split by percentile [13], as shown in Fig. 3. It is then expectable that each segment will have a different business value and that a user may want to target better valued business segments first.

Considering this idea, operations have been implemented in the Strategy Manager so as to allow the user to set sequences of business segments to be dialed. When defining the sequence, the user also defines which business goals control the transition from one business segment to the next.

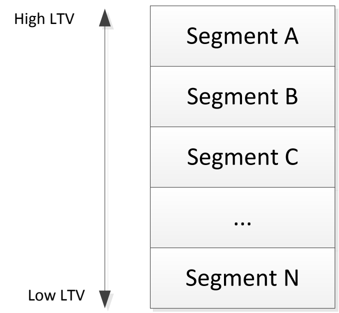


Figure 3. Business segment prioritization based on LTV.

G. Contact prioritization based on business value

Even after customers are segmented, they can still have a different individual expected business value. For example, each customer can have a different expected lifetime value, as suggested in [9]. So, it may be useful to contact customers with the higher expected value first (Fig. 4).

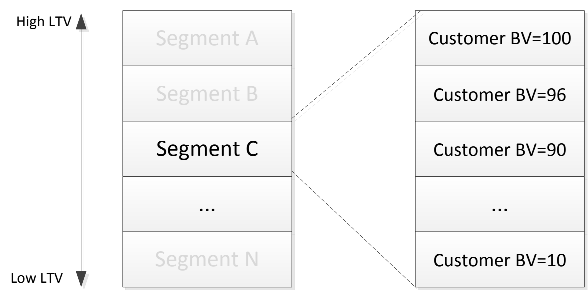


Figure 4. Contact prioritization according to business value.

In order to satisfy this requirement, the Strategy Center provides an operation to sort the contacts by a user-defined business value, which can be derived from each contact's attributes.

H. Multi-channel

As stated previously, contact centers are not restricted to telephony interactions and businesses find it important to be able to reach customers in multiple media [2][14]. Having this idea in mind, the Strategy Manager provides a mechanism to reach customers from a given segment through several media.

For example, suppose a contact center manager wants to define that, for "Business Segment X", after trying to reach a customer by phone by a maximum number of tries he should be reached by email instead. This behavior can be configured in Strategy Manager by defining a set of media rules for a given business segment.

 ISBN: 978-1-61208-325-4

V. Conclusion and Future Work

In this paper, we presented a module that implements customer segmentation for outbound contact center dialers with a focus on business intelligence. We presented a Contact Center Management System (Altitude uCI), where the module was added, and provided an overview of customer segmentation concepts. We have shown how it was possible to enhance an outbound dialer with customer segmentation concepts, focusing on schedules, KPIs, multi-dimensional statistics, business segment and customer contact prioritization by business value, and staffing. Some of our customers have been integrating their customer segmentation in the new "Strategy Manager" module, and their staff productivity and responsiveness to events regarding contact list quality has improved dramatically, as they can now monitor each segment's quality in terms of dialing and sales success and change dialing priorities and staffing accordingly and in real-time.

In the future, our research will be focused on enhancing the dialer with "best time to call a given customer" estimations, so that the probabilities of desired business outcomes are maximized. This may include classification modeling data mining techniques, possibly using decision trees (or other classification method) applied to the customer attributes, providing a classifier for call success or failure, in a given schedule [15]. In this scenario, the classifier also outputs a probability for "call success" for each customer. This probability output from the classifier can then be used to order customers by success likelihood.

Additionally, automatic customer segmentation mechanisms will be addressed. These will allow a customer list to be automatically segmented, with minimal user input. For example, automatically segmenting (binning) customers based on LTV intervals. The user would only select the field where to perform the binning [20]. Another possibility will be to use unsupervised machine learning techniques, to find related customer segments. Namely, K-means, Two Step, and Kohonen networks techniques can be used to automatically find customer segments. Additionally, a decision tree classification method could be applied to these segments, in order to find a human intelligible description for each cluster, which is also usable as a filter in the Strategy Manager business segments [20].

Adding more native support for marketing related customer segmentation types is also a future possibility. This would include support for segmentation types like Value-Based Segmentation, Behavioral Segmentation, Propensity-Based Segmentation, Loyalty Segmentation, Socio-demographic and Life-Stage Segmentation integrated into the system, automatically guiding the user into finding the right segments and providing the corresponding filters to the Strategy Manager [20] [8][9].

The addition of marketing related customer segmentation concepts and processes for specific industries is also a possibility. Banking [20], telecommunications, and electric power industry [10] could be considered, among others.

Better expressivity between business goals and KPIs may also be addressed, as proposed in the Business Intelligence Model research literature [16][17]. This may allow better business rules for prioritizing business segments.

Acknowledgment

This work was supported in part by QREN (*Quadro de Referência Estratégico Nacional*).

References

[1] N. Gans, G. Koole, and A. Mandelbaum, "Telephone call centers: tutorial, review, and research prospects", Manufacturing & Service Operations Management, Vol. 5, No. 2, March 2003, pp. 79-141.

[2] STRATECO, "Shaping tomorrow`s contact center: research on market trends & challenges - Germany" , http://online.altitude.com/germany-cc-market-survey-availability , April 2013.

[3] T. Weinberg, "New community rules:marketing in the social web", O'Reilly, ISBN: 9780596804237, 2009.

[4] P. Amaral, and M. Vital, "Predictive dialer intensity optimization using genetic alghoritms", International Journal of Machine Learning and Computing, in press, 2013.

[5] S. Chaudhuri, U. Dayal, and V. Narasayya, "An overview of business intelligence technology",Communications of the ACM, Vol.54, No. 8, August 2011, pp.88-98.

[6] H. Krause, and D. Arora, "Controlling-Kennzahlen - Key Performance Indicators: Zweisprachiges Handbuch Deutsch/Englisch - Bi-lingual Compendium German/English", Oldenbourg Wissensch.Vlg, ISBN-10: 348659690X, March 2010.

[7] J. Jonker, N. Piersma, and D. Poel, "Joint optimization of customer segmentation and marketing policy to maximize long-term profitability", Expert Systems with Applications,Vol. 27 , Issue 2, August 2004, pp. 159-168

[8] C. C. H. Chan, "Intelligent value-based customer segmentation method for campaign management: a case study of automobile retailer", Expert Systems with Applications, Vol. 34, Issue 4, May 2008, pp. 2754–2762.

[9] S. Kim, T. Jung, E. Suh, and H. Hwang, "Customer segmentation and strategy development based on customer lifetime value: a case study", Expert Systems with Applications, Vol. 31, Issue 1, July 2006, pp. 101-107.

[10] E. Carpaneto, G. Chicco, R. Napoli, and M. Scutariu, "Customer classification by means of harmonic representation of distinguishing features," Power Tech Conference Proceedings, IEEE Bologna, Vol. 3, June 2003, pp. 7.

[11] N. C. G. Campbell, and M. T. Cunningham, "Customer analysis for strategy development in industrial markets", Strategic Management Journal, Vol. 4, Issue 4, November 2006, pp. 369-380.

[12] R. Batt, "Strategic segmentation in frontline services: matching customers, employees, and human resource systems", International Journal of Human Resource Management, Vol. 11, No. 3, June 2000, pp. 540-561.

[13] F. R. Dwyer, "Customer lifetime valuation to support marketing decision making", Journal of Interactive Marketing, Vol.11, No.4, pp. 6-13 , Fall 1997.

[14] A. Payne, and P. Frow, "The role of multichannel integration in customer relationship management", Industrial Marketing Management, Vol.33, issue 6, August 2004, pp. 527–538.

[15] I. H. Witten, and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)", TBS, ASIN**:** B001DU6J6U, 2005.

[16] D. Barone, J. Mylopoulos, L. Jiang, and D. Amyot,”Business intelligence model, ver. 1.0. Technical Report CSRG-607”, University of Toronto, March 2010.
[17] D. Barone, L. Jiang, D. Amyot, and J. Mylopoulos, “Reasoning with key performance indicators”, Proceedings of the 4th IFIP WG8.1 Working Conference on the Practice of Enterprise Modeling (PoEM11), November 2-3 2011, pp. 82-96.
[18] G. R. Tirenni, "Allocation of marketing resources to optimize customer equity", Doctoral dissertation, University of St. Gallen, 2005.
[19] P. M. Silva, E. Pereira, and T. Alberto, "Scalable contact center management with data partitioning and data integration", unpublished, 2004.
[20] K. Tsiptsis, and A. Chorianopoulos,"Data mining techniques in CRM: inside customer segmentation", John Wiley & Sons, 2011.
[21] K. R. Kashwan, and C. M. Velu, "Customer segmentation using clustering and data mining techniques.", International Journal of Computer Theory & Engineering, Vol.5, Issue 6, December 2013, pp. 856-861.
[22] F. Hamka, H. Bouwman, M. De Reuver, and M. Kroesen, "Mobile customer segmentation based on smartphone measurement", Telematics and Informatics , Vol. 31, Issue 2, May 2014, pp. 220–227.
[23] P. Fennemore, A. I. Canhoto, and M. Clark, "An investigation of existing and emerging segmentation practices in online social media networks", Henley Centre for Customer Management, December 2011.
[24] A. I. Canhoto, M. Clark, and P. Fennemore, "Emerging segmentation practices in the age of the social customer", Journal of Strategic Marketing, Vol. 21, Issue 5, July 2013, pp. 413-428.
[25] A. Czyszczoń, and A. Zgrzywa, "Automatic customer segmentation for social CRM systems", Computer Networks, Springer Berlin Heidelberg, 2013. pp. 552-561.

Find a Book! Unpacking Customer Journeys at Academic Library

Alma L. Culén and Andrea A. Gasparini
Department of Informatics, Group for Design of Information Systems
University of Oslo
Oslo, Norway
almira@ifi.uio.no, a.a.gasparini@ub.uio.no

***Abstract*—Academic libraries are especially poised to serve knowledgeable and technologically advanced user population: students and researchers. The technological advances are dictating significant changes for academic libraries. This paper is concerned with building awareness within the library around the need to re-think its role in academic life, its use of technology and willingness to co-innovate with users. The paper reports from four workshops that aimed to explore existing and future services offered by the academic library. Library employees, students and researchers were participants in all workshops. The participants were first informed about service design and its tools, and then engaged in creating customer journeys, using service design cards. The set of cards used was an of-the-shelf product, modified for the purpose, introducing images specific to the library and allowing for rating of services in terms of their importance. The paper reports on our findings from these workshops. One interesting finding is that librarians still focus mostly on physical space and personal services, such as organization of courses in the library, while students and researchers almost exclusively think of digital services, related to literature they need for their work.**

Keywords—service design; service design cards; touch points; innovation; customer journeys; academic libraries.

I. INTRODUCTION

During the last decade, the Internet has been a game changer [1] for academic libraries. It created a challenge for academic libraries by providing access to articles anytime anywhere through, for example, Google Scholar and other open access publications sites catering to academic communities. In the past, the main issue with Internet access to academic literature has been the lack of credibility. In [2], credibility is considered as a multifaceted concept with two primary dimensions: expertise and trustworthiness. Academics have trusted for centuries the expertise of the library to provide good and credible information. Yet, the same are now trusting Wikipedia, Google Scholar, and similar, to provide them with credible academic information [3]. In addition to the Internet, the appearance of disruptive technologies, such as eBooks first, and tablets later [4]–[6], has posed further challenges. In combination with cloud computing, interested students and researchers are able to create their own collections of teaching and research materials, always at their fingertips. The libraries are practically forced to re-think their role in academic life, their use of technology and willingness to innovate.

An academic library is a place where serving academic community, cultivating, preserving and expanding knowledge is raison d'être. However, due to technological developments and changing habits of the academic communities, the services, as well as the ways of delivering them, are changing. The changes also imply the need and interest in ways of evaluating library services [7] and designing new ones. Looking through a variety of definitions and concepts regarding service and service design, see [8]–[10], we consider the following characteristics of a service to be useful also for discussing the library services: intangibility, heterogeneity, inseparability and perishability (IHIP). *Intangibility* is often cited as the most important distinction between tangible goods (products) and intangible services. For example, the help to a student by a librarian, in form of information, is intangible. *Heterogeneity* addresses the fact that services, even when the product obtained through the service is the same, for example, a book, is depending on different service providers and thus may be experienced in variety of ways. For instance, an experienced librarian may provide a different service and customer experience than a new librarian, when a customer inquires about a book. The experienced librarian may be able to offer similar titles, supplementary references etc. It is, thus, often difficult to achieve uniformity of the service delivery, a 'standard' service. *Inseparability* of service addresses the fact that it is impossible to separate the supply or production of the service from its consumption. The interaction between a provider and a customer in an act of offering/consuming a service may also be seen as an act of co-creation of the customer experience with the service, and thus, the customer may be identified as service co-producer [11]. *Perishability* of a service is addressed in the literature in multiple ways. Many consider a service to be something that happens in the moment and thus cannot be saved for later. For example, even though a student can borrow a book from the library for 4 weeks, the service takes place at the time of checking the book out. Alternatively, one may consider the service as ending at the time when the book is returned to the library. Similarly, in a new library database system, one may not be able to make certain inquires which were possible in the old database, and thus some services related to those may perish.

One fundamental attribute of services is that they have value only when they are used [12]. Other relevant attributes are trust, fast delivery (speed) and consistency of the service [13]. These attributes have a crucial impact on the customers' experience of a service, but do not have to be equally relevant for the provider. For example, Amazon has

 ISBN: 978-1-61208-325-4

built on trust, while McDonalds on the speed of the service. Services offered by a private and public sector differ in some important ways. In the public sector, the motivation to innovate services or to co-create them with users is often reduced, since the public sector services are actually intermediaries between the state (the actual service provider) and the user [14]. This makes it more difficult to influence improvement of existing, or development of new services [14]. It is more difficult for providers to understand and evaluate customer's experience of the service [15]. Finally, public sector customer services design may involve some paradoxes [16] that are difficult to resolve. Thus, working with services in public sector may be more challenging than working with services within the private sector.

In this paper, we examine how the academic library views service design and co-creation of services with users. To this end, we have organized four workshops with library employees, students and researchers. Part of the time during workshops was used to introduce concepts from service design, as well as methods and techniques used in service design. This content is presented, in its condensed form, in Section II. The remaining time was split equally between creating customer journeys in today's library and exploring future services. The main tool used to create customer journeys was a set of service design cards. The use of cards and card sorting is common in human computer interaction; see for example [17]. More on specific cards used is also provided in Section II. The paper reports on tool modification in order to collect more meaningful data from workshops, as well as insights gained and lessons learned.

The paper is structured as follows: Section II introduces design thinking, service design, customer journeys, touch points and service design cards. Section III describes our case and presents the workshops. The discussion is provided in Section IV, and it is followed by a conclusion and future work in Section V.

II. Service Design

This section presents, very concisely, the material used as a theoretical background during the four workshops that were conducted in order to initiate the envisioning and re-thinking process around services in the academic library.

A. *Design Thinking*

In contrast to analytical thinking in science, designers have developed another way of thinking, called design thinking. Design thinking involves building of new cognitive patterns to grasp multiple knowledge and multiple perspectives, related to the context at hand, that are to be synthesized and transformed into new products or services. It combines the empathy for the context of a problem, knowledge and understanding of others and designers' creativity in generation of insights gained around the problem. The entire process, including the translation of all insights towards solutions, often happens with stakeholders within the context of use. In practice, it is a method of finding solutions by going through certain stages, typically very similar to those of interaction design: formulate the problem, investigate it, brainstorm, make prototypes, chose one, implement and find out how well the solution solves the problem. Design thinking has also allowed designers to move from a post-production and branding place and become active participants in the making of new products and services [18].

Using design thinking in design of services offers a possibility to better meet customers needs [19], based on understanding of their behavior, motivations and other responses while interacting with services.

B. *Service Design*

It seems straightforward to define service design (SD) as a design of new services or re-design of existing ones. Design of services is not new; services have existed for millennia, but the recent popularity of service design may be attributed to design thinking approach to service design (see [18], [19]).

Service design differs from product design in the act of "doing" of the design [12]. Service design also differs from interaction design in that it uses more explorative ways to challenge the problem area, as opposed to interaction design with its more analytic approach [20]. Our understanding of service design is in line with that of Schneider:

"*Service design is an interdisciplinary approach that combines different methods and tools from various disciplines. It is a new way of thinking as opposed to a new stand-alone academic discipline. Service design is an evolving approach; this is particularly apparent in the fact that, as yet, there is no common definition or clearly articulated language of service design"*. [21]

Ideally, the service design teams should include all stakeholders related to the service context, as well as service designers, and other professionals, as needed for a specific project. The first step in the process of SD, an equivalent to defining a problem space in human-computer interaction, is an agreement on the context and interests. Different research methods such as ethnography, immersion, shadowing, sense-making methods such as mapping (including blue-prints, Giga maps and customer journeys), safaris, expert interviews and self-directed tools such as diaries are all part of the SD toolkit.

For the purposes of this paper, customer journeys and touch points are central.

C. *Customer Journeys and Touch-Points*

One of the most effective processes in service design is being able to visualize a service offered by an organization or a company using a tool called a *customer journey*. Koivisto explains:

"*Services are processes that happen over time, and this process includes several service moments. When all service moments are connected the customer journey is formed. The customer journey is formed both by the service provider's explicit action as well as by the customer's choices*", [22].

The 'service moments' Koivisto talks about are called *touch points.* Touch points, as stated above, comprise a customer journey and provide understanding of the service over time. They are thus a central aspect of service design [23], [24].

D. Service Design Cards

A good tool, helping to understand and address touch points in the initial stages of service development, is a set of service design cards. The card set that we chose was developed as part of AT-ONE method, a practitioner-based method for service-design, aiming towards maximization of the innovation potential at early stages of service design, see [23]. Clatworthy provides six different use contexts for the card set and evaluates the usefulness of cards in these contexts and in relation to their intended function. The cards were found to help with team building in cross-functional teams. Further, they were found to be helpful in assisting with the analysis and mapping of existing situations, generating ideas for new solutions or approaches, needs elicitation and facilitation of communication. In addition, Clatworthy says that the cards "*afford embodied communication and embodied cognitive processes",* [23].

III. THE CASE

As stated in the introduction, our goal was to re-think services offered by the academic library. The establishment of User-Driven Innovation project in the context of academic library at the University of Oslo approximately three years ago, started us on a research activity concerned with investigation and experimentation around users' involvement in innovation processes within the library. We have considered students' potential as innovators [25], as well as the living lab approach [26]. In this work, with students as innovators, we have found that images facilitated initial communication well, and they helped established common understanding of the problem area. Thus, our experience was similar to findings reported in [23]. The natural course of action was to buy the card set from [23], as shown in Fig. 1.

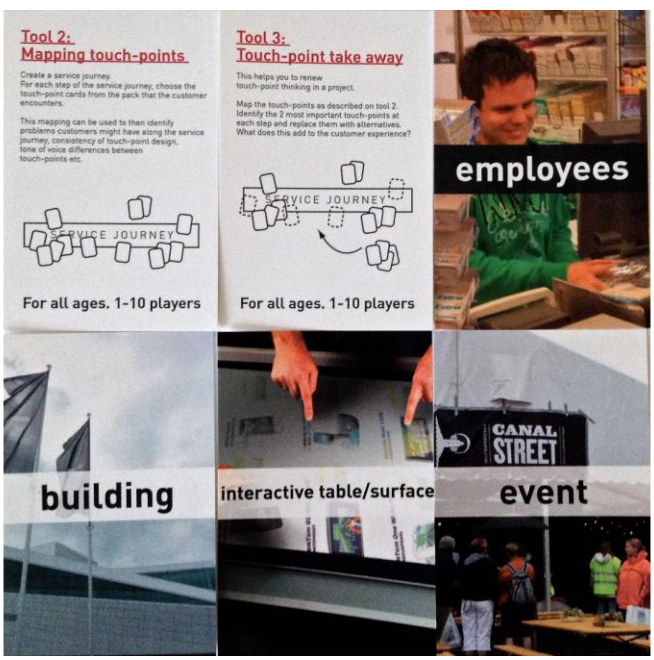


Figure 1. A sample of SD cards [23], including cards describing the two ways in which the cards were used in pilot workshops.

The set was then tested in couple of pilot workshops focusing on the library context and using the cards in two different ways: to map touch points and to remove a touch point from a customer journey. We chose to focus on customer journeys only, the removal of the touch point was deemed too specific. We quickly realized that some library specific cards would be helpful. This resulted in a modified set of cards which included vital touch points for the library, such as books, e-books, academic papers. Furthermore, we introduced two non-touch point cards, a critical point and a decision point, meant to be placed next to the touch point card in order to provide a clear visual clue related to the importance of the card. Then the set was tested one more time and we found out that differently colored dots placed next to the touch point card would be more useful in providing a graded importance clues. In addition, we found that colored arrows could provide further visual information, helpful in visualizing choices in the flow of the customer journey, again graded by importance. Thus, the set of cards used in workshops consisted of the original deck, plus added touch and non-touch point cards as described, and many dots and arrows in different colors, see Fig. 2, and Fig. 3 – Fig. 6, showing the cards in use.

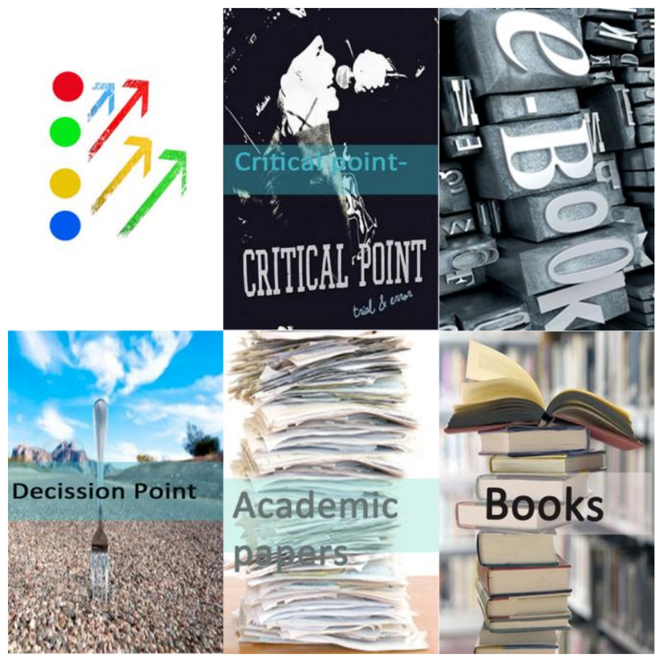


Figure 2. Cards suitable as touch points related to library services. Dots helped visualuize degree of relevance of a touch point, and arrows the flow.

When the cards were ready, we organized a series of four workshops, about 2 hours long, with similar set up. During the first hour, we introduced the concepts presented in the Section II: service design, design thinking, customer journeys, touch points and touch point cards. During the presentation two simple questions were asked in order to engage everyone in thinking about the library and innovation, and to invite the participants to be creative. The two questions were: "Can you give an example of a library service?" and "What does innovation in the library mean for you? Give an example."

During the second part of the workshop, service design cards were used to discuss a specific task. The task was to create a customer journey based on the following service provided by the library: find the literature relevant for a research or student project. When the journey was mapped, the new task was to envision this same journey in the future. During the first workshop, the same task was repeated using visual language in the making [27] for service design. This has not been done in other workshops.

The choice of the task was motivated by the sense of difficulty that users have when considering the role of the library in this particular process today (as explained in the Introduction, users often search Google Scholar and similar sites). It turned out to be a good choice for all participants. In fact, one of the researchers in the workshops admitted that she did not know that e-books purchased by the library are available for all university users, free of charge. The library employees could see that users did not have easily available information on this important new service, e-books access.

The same questions and the same task were used in subsequent workshops. While the first workshop involved many library employees, the remaining three workshops were predominantly composed of students, with at least one researcher and a library employee present.

Our main analytic tool was photo documentation [28]. A large number of photos were collected during workshops, so that we could analyze similarities and differences in processes with different groups, as well as how they made their journeys, for both present and the future service.

A. *The First Service Design Workshop*

The first workshop was held in May 2013, with 25 participants. 17 participants were library employees (included library leaders, librarians, subject librarians, digital services management, digital services support, e-resources consultants, open access consultant and others), 4 were students and 4 researchers. The participants were divided into four groups of six (seven in one of the groups) people, each group having at least four librarians, a researcher and a student. All four groups had their own deck of cards, dots and arrows, a large sheet of paper, and colored pens. The participants took some time to become familiar with cards, to discuss them and negotiate both the touch points and how to proceed with thinking about customer journeys. After 10-15 minutes, all groups decided on what touch points they would have on their journeys. Changes in the order of touch points and discussions became faster, as common understanding got established. Soon, all groups started using arrows and dots, Fig. 3. In one case, the paper under the cards was used to mark new paths between touch points that arrows could not reach. Also, some groups felt the need for additional card or two, or to document the process. Those were made using Post-it notes on the fly, and participants (mostly library employees and researchers) took pictures with their own mobile phones, showing engagement and importance of the subject discussed for their own work.

Looking at journeys made, we could notice that multiple starting points were deployed, usually from the physical to the digital. If the journey started in the digital world, it generally ended back in the physical world, in form of a visit to the library. This shows that meeting up at the library in person was regarded as vital in order to gain access to services.

The journey making allowed for relating anecdotes around how library services can be experienced by users. For example, in one of the groups, a student related a story of being charged a fee after the return of a book, which was long past due. For the library, the charge, as a source of income or compensation, is minimal and insignificant. For the customer, it provided for a really negative experience.

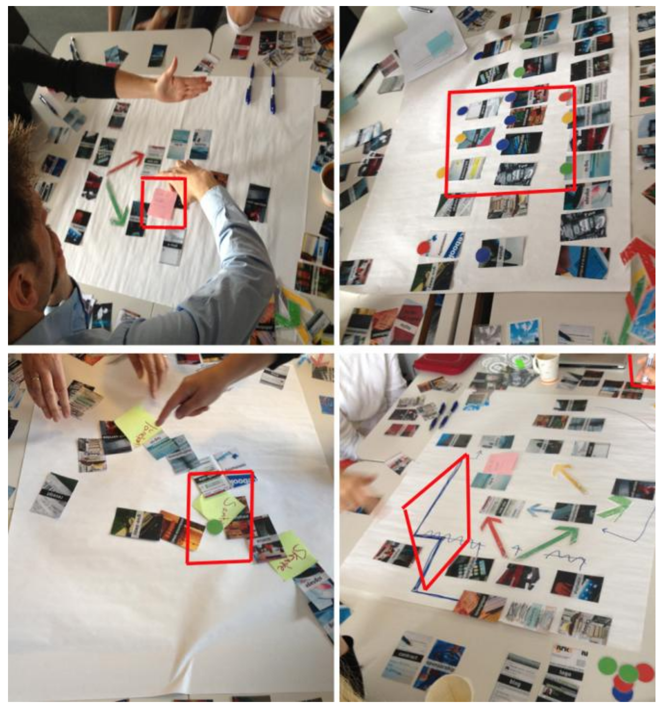


Figure 3. The workshop in the library. Each small picture shows the customer journey made by one of the four groups.

B. *The Second Service Design Workshop*

In this workshop, only three bachelor students from interaction design course participated, and one library employee/researcher. The same program, as described above for the first workshop, was followed.

The difference between the customer journey made at this workshop, and those that resulted from the first workshop, was where the service start points were, see Fig. 4. For the three participating students, it was not conceivable to start the journey elsewhere then with digital interfaces as touch points. The only reason that they added the physical library in the journey was because the library employee wanted to introduce it. The students added the card, and then quickly added a critical point card, from which an arrow lead to the library building card (Fig. 4), signaling clearly that only in times of absolute crisis would they venture into the building.

 ISBN: 978-1-61208-325-4

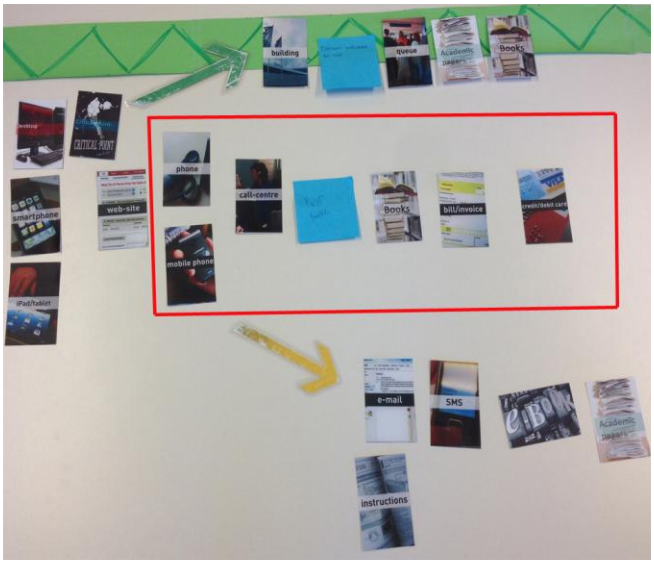


Figure 4. The journey always started from the digital: PC, a tablet or a phone.

Asked what kind of crisis they are thinking of, they exemplified with network failure, or the library site being down. An interesting outcome of envisioning the future service was the "book-to-door" service, a delivery of a physical book at home address, for which they were willing to pay.

C. The Third Service Design Workshop

The third workshop was carried out in the context of a graduate course in experimental design, with 18 participants divided into three groups. The journeys made are shown in Fig. 5.

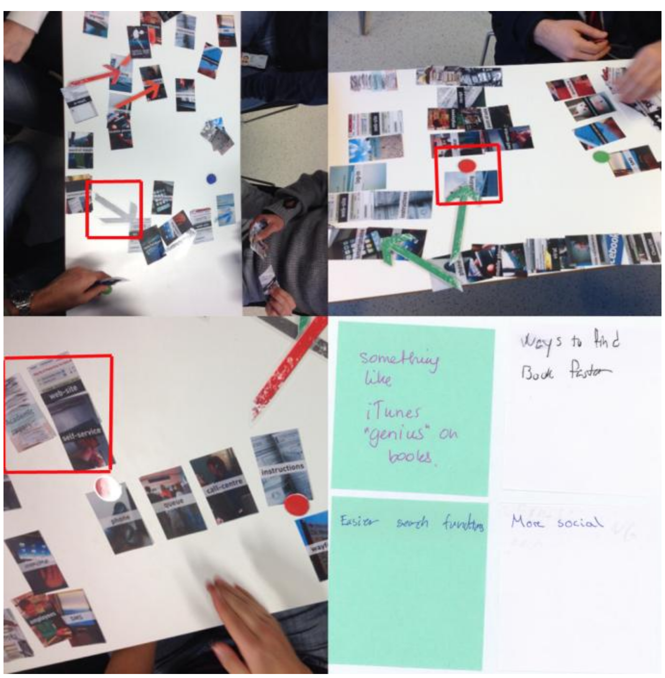


Figure 5. The three journeys and Post-it notes on new services.

In this case, too, all customer journeys (present and future) had starting points in the digital world. Similar to the previous group, as shown in Fig. 5, upper right image, the red dot was placed over the card depicting the building. Here too, the building became a touch point only in the case of an emergency. This group, however, would not opt for visiting the library at all, but would make a phone call instead.

Even though participants had ample time and seemed engaged in envisioning future services around finding academic literature, journeys they made remained quite conservative in terms of how far the ideas were from today's solutions.

D. The Fourth Service Design Workshop

The fourth workshop was also conducted with students taking a graduate course, but with ten students. The course had design and design thinking as a theme, thus little in terms of introduction was needed. We had expectation that these students would be more creative. This expectation was not met. The outcome was rather similar to journeys made in workshops two and three.

Figure 6. Students taking the course with focus on design and design thinking working on their journeys.

The use of the myths card is worth mentioning. A student related that she was afraid of visiting the library, since the library is a quite place, and she sees herself as being loud. She thinks that the library is not the right environment for her (see Fig. 6, the bottom right corner).

IV. DISCUSSION

After the last workshop with students, we felt that we have gotten much information on one hand, while on the other hand, for further development of ideas, a different team composition and format of workshops is needed. The workshops have shown that the card set worked well. It indeed engaged participants in discussions around touch points, and served as ground for building common understanding. They served well as a visual tool for understanding today's services and envisioning the future ones. The added cards worked well, and the participants considered the use of color-coded dots useful. In negotiating the touch points, groups sometimes added touch points that some members of the group were not convinced needed to be a part of the journey (like the library building in workshop 3). Placing a red dot on that touch point gave it a meaning (emergency only, for example). The color-coding was done by negotiating the meaning of the color within the group, and thus, was not the same across all groups. Arrows were used similarly. The participants were not encouraged to use either dots or arrows in any way, but most groups found them to be helpful in visualizing the journey.

Although all workshops were including library employees, researchers and students, the first workshop was the one where library employees accounted for a vast majority of participants. The outcome of that workshop was distinctly representing the library view of its services. It was very interesting to observe that most suggestions for innovation had to do with physical space and face-to-face communication, even if later had to be done using video. Fig. 7 shows a sample of what library employees gave as examples of innovation during the initial part of the workshop. As it can be seen, several suggestions had to do with avoiding queue in the library, while equally many were suggesting a video call canter. Later, while thinking about touch points and customer journeys, variation in envisioned journeys was greater. Yet, it was evident that the decision power in negotiations around customer journeys clearly belonged to the library employees.

On the other hand, the remaining workshops strongly represented users' point of view. The journeys, both present and future, show mostly digital touch points and digital services. Even when they were not digital, like the 'book-to-door' service from the second workshop, they did not involve the library building or face-to face communication. Even in cases of emergency, as in the example from the workshop three, the students seem to prefer other forms of contact, such as the phone, over the visit to the library.

Figure 7. Post-it notes with answers to the question "What does innovation in the library mean for you? Give an example".

From what we could hear and observe during the workshops, the perception of what a service should be characterized by (intangibility, heterogeneity, inseparability, perishability) also differs between the library employees and the library users. The students in the workshops were in favor of automated online services, in part, due to the equality in the delivery of the service and its independence of either the librarian, or the librarian's perception of the user. The customer journey in the second workshop (see Fig. 4, in the red box), describes the aforementioned book-to-door service, possibly inspired by Amazon [29]. Even though the service delivers a tangible object to the door, it still eliminates heterogeneity. The heterogeneity was not the characteristic that students valued. The librarians, on the other hand, consider their knowledge as extremely valuable, but highly dependent on the inseparability of the service. While the explicit part of their knowledge may become part of some system, the tacit knowledge and the long and rich experience in working with academic community is transferable only at the moment of service delivery in the physical world. If this experience and tacit knowledge are not valued enough by users, this kind of service is indeed in peril of perishing. From the outcomes of the workshops we can see that value and meaning of IHIP characteristics of services needs to be further negotiated among the stakeholders. The gap that exists needs to be bridged by design of services that better suit user ways of doing things, but do not lose the extremely valuable assets that the librarians have.

V. Conclusion and Future Work

The paper described the case of exploring the present and the future academic library services, focusing on finding literature relevant for some student or research project. Four workshops were organized. The first featured a large number of library employees, while the remaining three focused mostly on student population. The main tool for engaging stakeholders in the process of creating customer journeys was a set of service design cards, modified to best fit this very specific context. The process was photo documented.

First, we remark that that our extension of the service design cards was meaningful for our participants, as were the tasks. Several groups have explicitly and unsolicited mentioned that it was nice to be able to use the colored dots and arrows and 'grade' the importance of certain touch points. Only 2 groups have chosen not to use this possibility.

Our main finding shows that the library still focuses very much on physical services, while the students are nearly exclusively concerned with digital services. A very positive outcome of the whole process for the academic library is increased awareness of just how large this gap is and its recent commitment to engage in a long lasting user cantered project focusing on user experience and design of services with users.

The ideation and thoughts around new services that started in some of these workshops have already been carried a step further. One new service that helps find e-books has been designed and implemented. The future work involves participation in the afore mentioned project, which is now at its start, and deepening understanding of the role of human-computer interaction research, its methods and techniques, in design of user experience and services for the academic library.

ACKNOWLEDGMENT

The authors are indebted to all workshops participants, and the university library for initiating the project. The National Library, in part, financed the project.

REFERENCES

[1] M. E. Casey and L. C. Savastinuk, "Library 2.0: services for the next generation libraries," Library Journal. Vol.1(9), 2006, pp. 1–5.

[2] M. J. Metzger, "Making sense of credibility on the web: models for evaluating online information and recommendations for future research," J. Am. Soc. Inf. Sci. Technol., vol. 58, no. 13, 2007, pp. 2078–2091.

[3] F. J. Harris, I found it on the Internet: coming of age online. Chicago: American Library Association, 2011.

[4] A. L. Culén and A. Gasparini, "E-book reader and the necessity of divergence from the legacy of paper book," in Proceedings of the 4th International Conference on Advances in Computer Human Interaction, 2011, pp. 267–273.

[5] A. L. Culén, B. K. Engen, A. Gasparini, and J. Herstad, "The use of iPad in academic setting: ownership issues in relation to technology (non)adoption," Proceedings of the ICEM&SIIE, 2011, pp. 564–572.

[6] A. L. Culén and A. Gasparini, "iPad: a new classroom technology? A report from two pilot studies," in Information Sciences and e-Society, 2011, pp. 199–208.

[7] E. J. Davis, "What does success look like? Evidence-based assessment of the academic library", 17th Conference of Greek Academic Libraries: Academic library evaluation as a measure of In stitutional Quality Assessment, Ioannina, 2008, pp. 1–9.

[8] J.-V. Nisula, "Searching for definitions for service design - what do we mean with service design?", ServDes 2012, Co-Creating Services, 2012, pp. 171–175.

[9] C. Lovelock and E. Gummesson, "Whither services marketing? In search of a new paradigm and fresh perspectives," J. Serv. Res., vol. 7, no. 1, 2004, pp. 20–41.

[10] M. Barrett, E. Davidson, A.-L. Fayard, S. Vargo, and Y. Yoo, "Being innovative about service innovation: service, design and digitalization," International Conference on Information Systems (ICIS) 2012, Panel 7, 2012, pp. 1–5.

[11] S. L. Vargo and R. F. Lusch, "Evolving to a new dominant logic for marketing," Journal of marketing, 2004, pp. 1–17.

[12] A. Polaine, L. Løvlie, and B. Reason, Service design: from insight to implementation. Brooklyn, N.Y.: Rosenfeld Media, 2013.

[13] E. Fragnière, B. Nanchen, and M. Sitten, "Performing service design experiments using ethnomethodology and theatre-based reenactment: a swiss ski resort case study," Serv. Sci., vol. 4, no. 2, 2012, pp. 89–100.

[14] C.-K. Chen, C.-H. Yu, S.-J. Yang, and H.-C. Chang, "A customer-oriented service-enhancement system for the public sector," Manag. Serv. Qual., vol. 14, no. 5, 2004, pp. 414–425.

[15] A. Kumar, D. S. Lokku, and N. R. Zope, "Value driven approach for services design," Proceedings of the 56th Annu. Meet. ISSS, 2012, pp. 1–9.

[16] J. E. Fountain, "Paradoxes of public sector customer service," Governance, vol. 14, no. 1, 2001, pp. 55–73.

[17] G. Rugg and P. McGeorge, "The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts," Expert Systems, vol. 22, no. 3, 2005, pp. 94–107.

[18] T. Brown, "Design thinking," Harward Bus. Rev., vol. 86(6), 2008, pp. 84–95.

[19] R. Cooper, S. Junginger, and T. Lockwood, "Design thinking and design management: a research and practice perspective," Des. Manag. Rev., vol. 20, no. 2, 2009, pp. 46–55.

[20] S. Holmlid, "Interaction design and service design: expanding a comparison of design disciplines", Nordes, 2007, pp. 1–8.

[21] J. Schneider, M. Stickdorn, F. Bisset, K. Andrews, and A. Lawrence, This is service design thinking: basics, tools, cases. Bis B.V, 2012.

[22] M. Koivisto, "Frameworks for structuring services and customer experiences," in Designing Services with Innovative Methods, Finland, 2009, pp. 136–149.

[23] S. Clatworthy, "Service innovation through touch-points: development of an innovation toolkit for the first stages of new service development," Int. J. Des., vol. 5, no. 2, 2011, pp. 15–28.

[24] S. Miettinen and M. Koivisto (Eds.), Designing services with innovative methods. University of Art and Design Helsinki, 2009.

[25] A. L. Culén and A. A. Gasparini, "Student driven innovation: designing university library services," The Sixth International Conference on Advances in Human oriented and Personalized Mechanisms, Technologies, and Services, 2013, pp. 12–17.

[26] A. Gasparini and A. L. Culén, "University library as a living lab: innovating with students", The 2nd International M-sphere Conference for Multidisciplinarity in Science and Business, 2013, pp. 19.

[27] E. Lee and A. Karahasanović, "Can business process management benefit from service journey modelling language?," The Eighth International Conference on Software Engineering Advances, 2013, pp. 579–582.

[28] M. Crang and I. Cook, Doing ethnographies. London: SAGE, 2007.

[29] T. Treanor, "Amazon: Love Them? Hate Them? Let's Follow the Money," *Publ. Res. Q.*, vol. 26, no. 2, 2010, pp. 119–128.

3D Web-Based Shape Modelling: Building up an Adaptive Architecture

Ali Abdallah, Oleg Fryazinov, Valery Adzhiev and Alexander Pasko
The National Centre for Computer Animation,
Bournemouth University,
Poole, United Kingdom,
Email: ccengineer@gmail.com, {ofryazinov, vadzhiev, apasko}@bournemouth.ac.uk

***Abstract*—3D web-based shape modelling and rendering is becoming an increasingly important research area. Many applications have emerged, both established ones, such as collaborative design and new ones, such as heterogeneous objects modelling along with their subsequent fabrication using 3D printing. In this paper, we explore a crucial issue of the technology, which is building a proper adaptive architecture for an interactive client-server based system with a particular emphasis on rendering aspects. We identify the most probable scenarios of executing modelling and rendering in terms of server-client communications and associated decision making, and then describe a number of case-studies, which allowed us to experiment with different rendering techniques in the context of various task distribution and communications mode of the adaptive interactive client-server based system. Finally, we present and analyse the results and suggest a number of practical recommendations. The main results of the paper are concerned with consideration of rendering.**

Keywords-Adaptive architecture, 3D shape modelling, WebGL, Collaborative shape modelling, Function Representation

I. Introduction

3D web-based shape modelling and rendering is an important emerging area of recent research, especially in the context of collaborative computer aided design, distributed computer games, scientific visualization applications based on data from a number of sources, and other applications, such as increasingly popular remote fabrication and 3D printing on demand. A number of efficient software tools have been developed, which allow for a wider scope of web applications. The web browsers make use of languages, such as VRML [1], WebGL [2], etc., to describe complex 3D scenes.

Despite rapid development of hardware including specialised Graphics processing units (GPUs) and widening Internet bandwidth, truly interactive applications allowing for near real-time rendering without loss of a visual quality are yet to become a reality. Tools outputting models in the standard format of polygonal meshes have many drawbacks and limitations (especially in a collaborative mode), and inefficiency of communications through the network and rendering, thus being at odds with the current cutting edge of the technology in terms of hardware and networking abilities.

More specifically, in terms of building a proper client-server architecture for 3D web-based systems, the development of the collaborative tools for modelling and rendering requires a flexible and efficient handling of hardware and software resources of each client to achieve an efficient workflow. In this paper, we explore in depth the problems of building such an adaptive environment with a particular emphasis on rendering aspects, thus defining the most efficient way the content is transferred between the server and the clients depending on the available resources.

From the geometric point of view, the most common format to represent and store 3-dimensional objects is polygonal meshes that embody the Boundary Representation (BRep). This format is widely supported by the modern 3D APIs implemented in web browsers and quite a few methods to create and process content in this format exist. However, the polygonal meshes are by definition an approximation of the mathematically precise model and have well-known issues concerned with the loss of the shape precision and visual property definition, limited complexity, large memory consumption, problems with transferring through networks, etc [3]. These result in problems with interactivity which are so important in collaborative web-based projects. An inability to access the construction history is also an issue [4].

These problems have become increasingly critical in the context of modelling heterogeneous objects where their internal structure is to be represented and rendered. One of the solutions to this fundamental problem is using the Function Representation (FRep) instead or together with BRep, where a 3D object is represented by a continuous function of point coordinates (implicit surfaces are a particular case of FRep). It allows for dealing with objects as volumes with an internal structure, keeps the constructive tree of their modelling process and is compact by its very nature. However, the main problem with this representation is a need to convert an FRep model to existing supported representations during a rendering step which can be an expensive procedure with some difficult technical issues. In this paper, we focus on the rendering aspects of collaborative modelling systems with FRep at its core in the context of building the most efficient client-server network architecture for an adaptive environment.

In this paper, we explore different ways of building the interactive architecture for the modelling system based on FRep with different types of adaptations to the client needs with a particular consideration of rendering. The paper structure is as follows. After a survey of related works, we describe a 3D shape modelling architecture of Web-bases system considering in detail first a pure client-server one and then an adaptive one. Having identified four the most probable scenarios of executing modelling and rendering in terms of server-client communications and decision making, we then consider a number of case-studies, where different rendering techniques (based on Marching cubes, Hybrid WebGL and server based rendering by using C++) are used depending on the different client abilities. Finally, we present the results of the experiments with both simple and more complex models as well as some practical recommendations reflecting the advantages and drawbacks of the tested techniques.

II. Related Work

The related works include Web-based scientific visualization, collaborative shape modelling, Computer-Aided Design and Manifacturing (CAD/CAM), and Web 3D concerned with delivery and interaction with 3D geometric models on the Web.

The overview of methods allowing to handle visualization-specific representation that consists of registered and merged points, surface and volume data as well as the corresponding meta information in order to provide important features for scientific visualization was presented in [5]. The survey shows that the information is usually re-sampled onto a structured regular grid after data acquisition and before filtering the data accordingly. However, transmission of vast data arrays prevents interactive modifications and collaborative work.

Most of the tools allowing collaborative solid modelling adopt Boundary Representation as the data exchange format defining an object by a set of surface patches stitched together. Thus, Kao proposed a collaborative CAD/CAM system CO-CADCAM that includes surface modelling, tool path simulation and post-processing, and CAD geometry co-editing [6]. Another example is a NURBS modelling system that supports a multi-user environment for collaborative conceptual shape design [7]. Sharing and editing of a solid model over the web can be done by collaborative solid modelling, but requires a whole modeller installed at each client [8]. Some collaborative modelling systems restrict the editing to the limited set of operations, such as modification of specific features of the model [9].

In web-based modelling, the main question is interaction between server and the client, i.e., the information that server sends to the client to render the model. In case of volume data the amount of information transmitted between a server and a client can be significantly large and the client is often requires to install an additional software tool. Thus, in X3D format [10], 3D objects supporting point, surface and volume primitives are described, but additional plugins for the browser needs to be installed. Web-based direct volume rendering with ray-casting was discussed by Congote et al. in [11] for the purposes of medical imaging and radar meteorology. Another way to send the volume information is Bidirectional Texture function that allows the progressive transmission and interactive rendering of digitized artifacts consisting of 3D geometry and reflectance information [12]. The similar approach is used in X3DOM where a lightweight geometry is compressed and transmitted with so-called image geometries [13].

Service-Oriented Architecture (SOA) [14] composes several low-level services to more complex services with a higher level of abstraction. SOA utilizes REST, JSON and XML-based web-service protocols and helps in supporting the collaboration between different applications running on different platforms as discussed in [14]. Koller et al. transfer images to the client and include a number of active defense methods to guard against 3D reconstruction attack by providing an interesting proposal to the protection system with a remote rendering service [15].

FRep based experimental systems for interactive and collaborative modelling on the Web include HyperFun Java applet [16], EmpiricalHyperFun [17], FVRML/FX3D [18], XISL [19], Hyperfox plug-in to Firefox [4], and a BlobTree implementation with websockets [20]. Most of these works had one solution implemented for rendering, mainly the isosurface polygonizaion on the client side. Today without the need for any plug-in installation process, WebGL provides access to the native GPU layer for rendering in a browser on the client side, which provides a basis for adaptive rendering architectures.

III. Adaptive 3D Shape Modelling Architecture

As we discussed above, our main motivation in using FRep models is the reduced complexity of the models allowing to avoid well-known problem of handling large 3D data files and vast amount of computational resources. However, the major drawback of these models is difficulties of handling them inside a web-browser because of the lack of the native support of non-polygonal objects in the current standards for rendering 3D scenes. Modern browsers allow only to load 3D models defined in the form of polygonal mesh and interactively manipulate these scenes (translate, rotate, scale) inside web browser with an input device, such as mouse or keyboard. Usually, the conversion between an FRep model and a polygonal model is required with the process called polygonization. In this section we discuss different ways of implementing web-based shape modelling with FRep objects.

A. Pure Client-Server Architecture

The client/server architecture can be considered as a network environment that exchanges information between a server machine and a client machine where server is a large-capacity computer, with a large amount of data stored on it and available for sharing with different clients. The clients are smaller computers that are used to perform local computer tasks. The client/server architecture reduces multiple copying of a single file and allows an organization to have one centralized point for every computer to access the same application.

The system we discuss here is platform-independent from the client point of view. As the client can have very small abilities to handle geometric data, we consider the server to be responsible for performing most of the computationally intensive tasks. By computationally expensive tasks, we consider primarily the point queries, i.e., calculation of the function value for the given point in space. The point query is mostly used for rendering purposes. It can be shown that all the other operations in a shape modelling system based on FRep, such as modification of the function, adding primitives and operations to the defining function and others, are very cheap and can be done on virtually each client. Therefore, in the pure client/server architecture the user still can create objects and define operations, import and export objects in the appropriate file format, or import geometry in common file formats.

The proposed three-tier architecture encompasses a client-side, a server-side and a database. The client side has a copy of the master 3D model, rendered in the web browser with one of the techniques discussed below. The web browser GUI contains a number of tools for creating primitives and performing operations on the model with using user's hardware resources (GPU and local memory). The server side contains the master copy of the model with all the references to the external files stored in the database and the kernel modelling system. The server is responsible, as mentioned above, for performing most of the computationally expensive tasks. The task of transferring data through a network medium, such as the Internet or an intranet, is performed by a communication layer between the clients and the server.

Depending on the abilities of the client, the data can be transferred from the server to the client as either polygonal mesh objects, WebGL texture objects or as pure image files. In the first two cases the picture is generated with WebGL, in the third one - by native browser's image handling. The work on the server and on the client is connected by a code written in JavaScript. In the case of WebGL rendering, the client side utilizes X3DOM, which uses JavaScript with WebGL.

We would like to stress that the resulting system should be scalable meaning that the client can be either desktop or mobile with different hardware abilities. For example, mobile clients can often process only image data and therefore the server should be able to stream images instead of 3D data.

B. Adaptive Architecture

As we mentioned before, the hardware on the client side can be very different. Different parameters of the client should be taken into account to choose the best possible way to deliver the rendered model from the server to the client. These parameters include: type of the client machine (desktop or mobile), CPU and GPU availability and power, amount of CPU and GPU memory and the supported software (such as WebGL support in the browser). Therefore, an adaptive architecture is the one that takes all the parameters into account to process the objects from the server to choose the best possible way to deliver model rendering to the client.

From the rendering point of view, the FRep model can be rendered in two different ways:

- Polygonization, i.e., conversion of an FRep object surface into a polygonal mesh and then rendering of the resulting mesh by traditional tools;
- Direct rendering, usually in the form of ray-casting.

In the client-server chain, the client machine potentially can be more computationally powerful than the server and in this case the adaptive architecture should take that in mind and transfer not the result of the rendering to the client but rather the model itself, such that the client performs the rest of the rendering tasks. If rendering takes part on the server, different ways to send the result to the client can be used. These include delivery of the result as

- images obtained after the direct rendering;

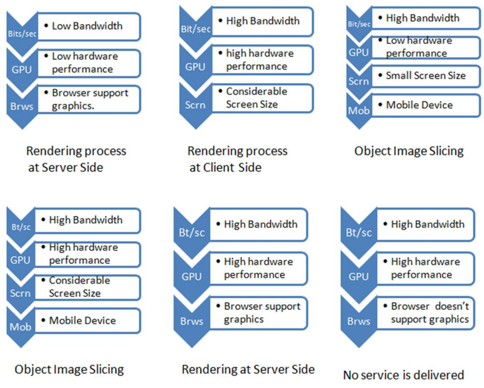


Figure 1. Different Types of Scenarios and Decision Making.

- objects delivered as image slices (voxel array);
- objects delivered as discrete data structures, for example point clouds or polygonal meshes;

Therefore, the adaptive shape modelling architecture or environment is the one that is able to adapt or react and interact with clients according to the abilities of the client, and uses different scenarios depending on the needs and the abilities of the client by choose the most efficient way.

To make a decision regarding the appropriate scenario, the adaptive architecture gets the information about the client machine while communicating in the background to retrieve the data about the characteristics of the hardware and the browser and to calculate the bandwidth download rate. The retrieved information helps the environment to analyze each client or user and decide which scenario it should use for this particular client. It can be seen that lots of various combinations of the types of rendering, machine for rendering and others can lead to large number of different combinations (Fig. 1). However, in practice we identified four most common scenarios. They are the following:

1) The request from the server identifies a client with low bandwidth and low computational resources, the server renders the model and sends the result as a low-resolution static image.
2) The client has sufficient bandwidth, but the browser is not capable to render 3D object, the server renders the model from different points of view and sends the result as set of mid-resolution static images to allow transformation of the object on the client.
3) The client has sufficient bandwidth and is capable to render 3D objects, however, it has low computational resources comparing with the server. The server performs polygonization of the model and sends the resulting polygonal mesh to the client. The client renders the polygonal model using its own resources.
4) The server identifies the high performance machine on the client side, it sends the complete model to the client, such that the client can render the model using local resources. In this scenario, the server performs only the model data transfer, because no rendering takes place on its hardware resources.

In the following section, we present implementation and experiments with some of the identified scenarios of FRep models rendering on adaptive architectures.

IV. Implementation and Results

Leaving questions of collaborative editing of FRep objects within a shape modelling system beyond the scope of this paper, we want to focus more on the rendering aspects of such a system. The core of this system is the adaptive architecture we described above. For our experiments, we implemented a prototype of this adaptive environment that allows to work with pre-defined FRep models of different complexity. Below we discuss some aspects of the prototype and the experimental results.

A. Adaptive architecture implementation

Being a web-oriented system, most of the tasks for adaptation take part in the browser, meaning in the JavaScript code. Thus, at the start of the work, the server requests the information about the client by running specialised JavaScript module. The client sends the requested information (connection speed, machine info, browser info detection, OS info, screen resolution info) to the server as XML messages. Of course, there is no need to re-send information, such, as OS info and IP address, that is not going to change. The server selects one of the above mentioned scenarios based on the information requested from the client. This process is periodically repeated during the work while client is connected to the server to ensure that the selected scenario needs to be continued.

As it can be seen from the scenarios we discussed in the previous section, most critical information that server gets from the client is the connection speed and the browser ability to use WebGL. The connection speed detection begins when the client starts downloading a certain file with predefined size in kilobytes from the server in the background; a timer starts as soon as the downloading process starts, the duration is obtained as soon as the download process stops, the transfer bit rate is then calculated by dividing the size of the file by the duration. The duration is calculated by subtracting the end time from the start tame and dividing the result by one thousand (to convert from milliseconds to seconds), the number of bits loaded is calculated by multiplying the downloaded file size by eight, the speed of download is obtained by dividing the number of bits loaded by the duration obtained. The diagram (Fig. 2) describes the machine info detection process starting by determining the agent of the user and check whether it is a portable device or not. Determining other mobile devices can be performed by the operating system detection process. Detection of WebGL abilities can be done either from the canvas initialisation, meaning that successful initialisation indicates that the machine supports WebGL, or by analysing web browser information. For example, information that the browser is Internet Explorer of version 10 and early means that WebGL is not supported.

B. Rendering techniques comparison

For our tests, we implemented three rendering techniques corresponding to three different scenarios. All the tests were done on the same machine running Chrome web browser with high-speed network connection to the server. The first technique is rendering on a server (Fig. 5) by using C++ code and sending still images to a client. The second one is polygonization on the server and rendering the resulting mesh on a client by using WebGL (three.js library) (Fig. 4). The third one is performing polygonization on a client by applying marching cubes technique implemented as JavaScript code (Fig. 3). Note that in the second scenario the polygonization is done with C++ code on a server while in the third scenario the polygon mesh is being created on a client side implemented in JavaScript.

Figure 2. 3D simple objects created using marching cubes, with low quality (resolution) and sharp edges.

Figure 3. 3D simple objects rendered using WebGL library (Three.js) for rendering, objects are with smooth edges and higher resolution.

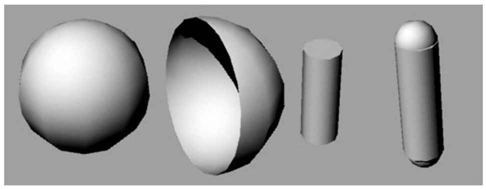


Figure 4. 3D simple objects rendered at the server side (using C compiler) and sent to the client as images.

Models with different complexity were used in our tests (Figs. 5,7,8). We show the timings of the rendering by using different methods in the Table 1. To achieve similar performance rate, low resolution was used for rendering on a client side that resulted in visible edges and lower quality of the result (Fig. 3). The resolution can be increased, but in this case more computational resources from the client are required.

It can be seen that for simple models different rendering techniques showed no major difference in timings. However, as the complexity of the object increased, the difference in timings becomes more clear. In general for the same initial conditions, the best timings were achieved in case of rendering purely on a server side and sending the result as an image to the client. At the same time it can be seen that the ability to work with model in interactive way is limited, as we have to re-render the model for each new camera position.

The chart in Fig. 9 shows the performance for the three rendering techniques with respect to shape models of different complexity. The green curve represents the 3D objects rendered using C++ at the server side. It shows that complex

 ISBN: 978-1-61208-325-4

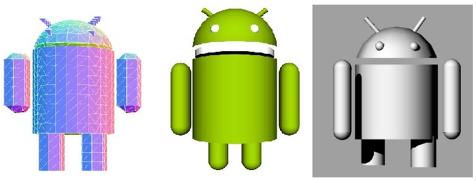


Figure 5. Android Robots redered using different techniques, the robot to the left was created using marching cubes, the robot in the middle was rendered using WebGL library (Three.js) and the robot to the right was rendered at the server side using rhinoceros and sent as image to the client.

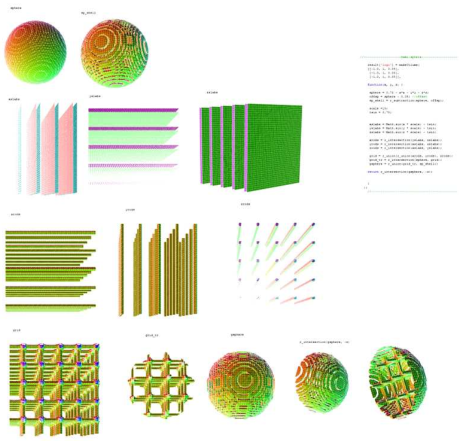


Figure 6. Image slicing showing the different rendering phases the complex Hemi-sphere model took after applying different functions using marching cubes.

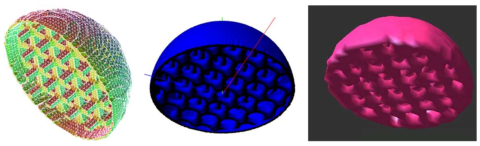


Figure 7. Hemi-Sphere models (Complex Models) were rendered using different techniques, the complex model to the left was created using marching cubes, the one in the middle was rendered using WebGL library (Three.js) and the one to the right was rendered at the server side using C Language (C++ in our case) and sent as image to the client.

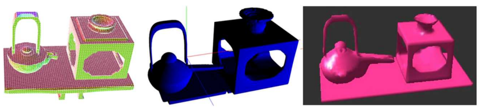


Figure 8. Sake pot models (Complex Models) were created using different techniques, the complex model to the left was created using marching cubes, the one in the middle was rendered using WebGL library (Three.js)and the one to the right was rendered at the server side using C++ and sent as image to the client.

objects (Hemi-sphere and Sake Pot models) were rendered in a fraction of a second. The blue curve represents objects rendered using polygonization with the Marching Cubes. The rendering time stayed almost constant when rendering simple objects, and rose when rendering complex objects. The rendering time for the hemi-sphere and the sake pot objects was less than half a second. The red curve representing rendering with hybrid WebGL started to rise up sharply when rendering complex objects, this indicates the amount of time and the

TABLE I. COMPARING DIFFERENT RENDERING TECHNIQUES(MARCHING CUBES, WEBGL USING THREE.JS, AND SERVER SIDE RENDERING), THE TABLE BELOW SHOWS THE TIME NEEDED IN MILLISECONDS FOR EACH TECHNIQUE IN ORDER TO CREATE A SIMPLE 3D OBJECT

3D Object	Marching Cubes	WebGLThree.js	Server Rendering
Sphere	0.006	0.003	0.0013
Semi-Sphere	0.005	0.03	0.0013
Cylinder	0.003	0.02	0.0012
Closed-Cylinder	0.007	0.025	0.0013
Android Robot	0.021	0.196	0.178
Hemi-Sphere	0.168	1.782	0.0012
Sake Pot	0.507	4.597	0.0018

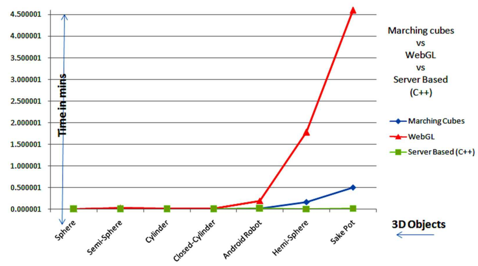


Figure 9. Chart diagram shows the time taken by each 3D model when applying different rendering techniques.

power of GPU needed to perform such rendering,

As a result, the objects rendered with Marching Cubes are ten times lesser in rendering time and GPU power, and it is perfect for users with low GPU performance, and/or low Internet bandwidth, while clients with high GPU performance can rely on hybrid WebGL. Users with low GPU performance and high Internet bandwidth can use server side rendering where objects with high resolution can be rendered (Fig. 6).

The bar chart in Fig. 10 compares the rendering time taken to obtain the same object using the three rendering techniques, it is clear that the hybrid WebGL rendering method came in the first place in terms of GPU high performance, and high rendering time, the Marching Cubes came second with considerable performance on the GPU and rendering time, and last came the server side rendering method using C++, which is very efficient in both GPU power and timing.

V. CONCLUSIONS AND DISCUSSION

In this paper, we have considered the features of a Client-Server adaptive architecture to establish the most convenient and efficient way of 3D Web-based modelling and rendering with a particular emphasis on the latter. We started from outlining some advantages of using the Function Representation over polygon meshes for modelling system; then we described the specifics of an interactive client-server architecture for modelling and rendering, and identified four most common scenarios of executing those processes along with the necessary communication and decision making using that adaptive architecture.

The main results of the paper are concerned with consideration of rendering. To explore the specifics of the proposed architecture with varying client and server abilities in the context of the outlined scenarios we have implemented three

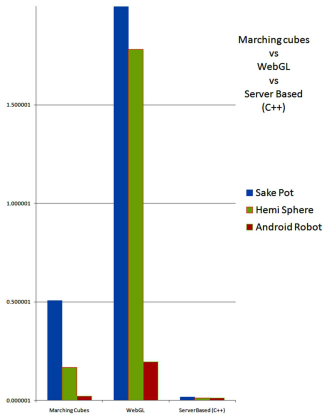


Figure 10. Bar chart diagram comparing the rendering time taken by each complex model when rendered using three different techniques.

different rendering techniques, namely: Server-based rendering using C++ code (with sending still images to the client); polygonization on the server and rendering the resulting mesh on a client using WebGL; Marching Cubes based polygonization implemented as JavaScript code on the client side. First, we showed as a prove of concept that rather simple objects can successfully be rendered using different techniques. Then, two more complex models (Hemi-sphere and Sake pot) were rendered using the three identified methods. The detailed comparative numerical and visual results have allowed us to make the following conclusions:

1) Gathering information about a client requesting the service is a key factor in determining what kind of service (i.e., rendering technique) to deliver
2) Different scenarios should be considered, the proposed architecture allows for reliable and efficient rendering process
3) Different rendering techniques can be applied over objects of various degree of complexity, although the required GPU power and the necessary rendering time can vary in a significant extent for simple and complex models.
4) Rendering at the server side using C++ implementation can be very efficient in terms of processing power and rendering time; however, the problem is that objects are delivered to clients as still images
5) Rendering using Marching Cubes is very efficient in both rendering time and GPU power as it is done at the client side; however, this method properly works only for low resolution objects.
6) Rendering using hybrid WebGL and Marching Cubes techniques allows for high resolution objects and proved to be the optimal solution on machines with high GPU power.
7) Building up an adaptive environment, which is capable to interact and deal with different kinds of users is a challenge, and still needs further research.

The following recommendations reflecting the advantages and drawbacks of the tested rendering techniques provided by the adaptive environments can be stated:

- More experiments on more complex objects should be executed to further analyze the behavior of these objects in terms of their modelling and rendering within the proposed environment.
- Other rendering techniques, such as ray casting and volume rendering should be investigated.
- Implementation of an intelligent engine in the core of the adaptive environment promises more functionality for making decisions and thus for providing better rendering services.

Future work will include exploration of collaborative web-based modelling and rendering of heterogeneous objects with a complex internal structure in the context of a flexible interactive adaptive architecture including testing different scenarios of modelling and rendering with taking into consideration different multiple platform configurations.

REFERENCES

[1] R. Carey and G. Bell, The annotated VRML 2.0 reference manual. Addison-Wesley Longman Ltd., 1997.
[2] “OpenGL ES 2.0 for the web,” http://www.khronos.org/webgl/.
[3] R. Balan and G. Taubin, “3d geometry compression and progressive transmission,” Computer-Aided Design, no. 32, 2000, pp. 825–846.
[4] T. Vilbrandt, O. Fryazinov, C. Stamm, and A. Pasko, “A web oriented function-based volume modeling framework,” Computer Graphics & Geometry, vol. 12, 2010, pp. 41–51.
[5] R. Bürger and H. Hauser, “Visualization of multi-variate scientific data,” in EuroGraphics 2007 State of the Art Reports (STARs), 2007, pp. 117–134.
[6] Y.-C. Kao and G. C. Lin, “Development of a collaborative cad/cam system,” Robotics and Computer-Integrated Manufacturing, vol. 14, no. 1, 1998, pp. 55 – 68.
[7] C. V. Foster, Y. Shapirstein, C. D. Cera, and W. C. Regli, “Multi-user modeling of nurbs-based objects,” in ASME Design Engineering Technical Conferences, Computers and Information in Engineering Conference (DETC 2001/CIE-21256), ASME, 2001.
[8] S. Chan, M. Wong, and V. Ng, “Collaborative solid modeling on the www,” in Proceedings of the 1999 ACM Symposium on Applied Computing, ser. SAC ’99. New York, NY, USA: ACM, 1999, pp. 598–602.
[9] R. Bidarra, E. Berg, and W. F. Bronsvoort, “Interactive facilities for collaborative feature modeling on the web,” in Proc. of the Tenth Portuguese Conference on Computer Graphics, 2001, pp. 43–52.
[10] “X3D - extensible 3D, New-Generation Open Web3D Standard,” http://www.web3d.org/x3d/.
[11] J. Congote, A. Segura, L. Kabongo, A. Moreno, J. Posada, and O. Ruiz, “Interactive visualization of volumetric data with webgl in real-time,” in Proceedings of the 16th International Conference on 3D Web Technology, ser. Web3D ’11. ACM, 2011, pp. 137–146.
[12] C. Schwartz, R. Ruiters, M. Weinmann, and R. Klein, “Webgl-based streaming and presentation framework for bidirectional texture functions,” in The 12th International Symposium on Virtual Reality, Archeology and Cultural Heritage VAST 2011, Eurographics Association. Eurographics Association, Oct. 2011, pp. 113–120, best Paper Award.
[13] J. Behr, Y. Jung, J. Keil, T. Drevensek, M. Zoellner, P. Eschler, and D. Fellner, “A scalable architecture for the html5/x3d integration model x3dom,” in Proceedings of the 15th International Conference on Web 3D Technology, ser. Web3D ’10. New York, NY, USA: ACM, 2010, pp. 185–194.
[14] T. Erl, Service-Oriented Architecture: Concepts, Technology, and Design. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2005.
[15] D. Koller, M. Turitzin, M. Levoy, M. Tarini, G. Croccia, P. Cignoni, and R. Scopigno, “Protected interactive 3d graphics via remote rendering,” ACM Trans. Graph., vol. 23, no. 3, Aug. 2004, pp. 695–703.

[16] P.-A. Fayolle, B. Schmitt, Y. Goto, and A. Pasko, "Web-based constructive shape modeling using real distance functions," IEICE - Trans. Inf. Syst., vol. E88-D, no. 5, May 2005, pp. 828–835.

[17] R. Cartwright, V. Adzhiev, A. A. Pasko, Y. Goto, and T. L. Kunii, "Web-based shape modeling with hyperfun," IEEE Computer Graphics and Applications, vol. 25, no. 2, 2005, pp. 60–69.

[18] Q. Liu and A. Sourin, "Function-defined shape metamorphoses in visual cyberworlds," Vis. Comput., vol. 22, no. 12, Nov. 2006, pp. 977–990.

[19] J. Parulek, P. Novotny, and M. Sramek, "Xisla development tool for construction of implicit surfaces," in In SCCG 06: Proceedings of the 22nd spring conference on Computer graphics, Comenius University, Bratislava, 2006, pp. 128–135.

[20] H. Grasberger, P. Shirazian, B. Wyvill, and S. Greenberg, "A data-efficient collaborative modelling method using websockets and the blob-tree for over-the air networks," in Proceedings of the 18th International Conference on 3D Web Technology, ser. Web3D '13. New York, NY, USA: ACM, 2013, pp. 29–37.

Methodology for Designing User Test Environments to Evaluate Web Accessibility Barriers with Disabled Users

Afra Pascual, Toni Granollers
GRIHO research group
Polytechnic School,
University of Lleida (UdL)
Lleida, Spain
apascual, tonig @diei.udl.cat

Mireia Ribera
Adaptabit research group
Department of Information Science,
University of Barcelona (UB)
Barcelona, Spain
ribera@ub.edu

Jordi L. Coiduras
Competect research group
Department of Pedagogy and Psychology
University of Lleida (UdL)
Lleida, Spain
coiduras@pip.udl.cat

***Abstract* — This work presents a new methodology for designing user tests that evaluate the impact of different Web accessibility barriers affecting people with disabilities. The methodology is based on several steps that help in the creation of Web content to be tested by users with specific disabilities. Several user tests have been carried out to check the validity of the steps defined in the methodology. The paper includes the methodology used, a case study and the key findings of the analysis. The results of the case study have been positive as the objective of providing a methodology has facilitated the creation of an evaluation environment: selection of the elements to be evaluated and the tasks to be performed by users who carried out the test. As a conclusion, we can state that the introduced methodology helps to include elements in user tests and optimizes the time and available resources needed for the preparation of an accessibility test.**

Keywords - Design, User test, Barriers, Web accessibility, Users with disabilities

I. INTRODUCTION

A large number of people have difficulty accessing the Web. The World Report on Disability [1] estimates that between 15% and 19% of the population have some type of disability, a substantial group of consumers that should not be ignored. However, current accessibility involves a compromise between ideal accessibility and available resources. As in classical usability, it is common sense to place more emphasis on problems with greater impact and/or that recur more frequently [2]. Therefore, it is important to understand the difficulties caused by different elements when experienced by specific user groups, in order to prioritize actions taken to improve accessibility for them. Technical accessibility, the most widespread, attempts to place all users and all contexts on the same level, but usability –and accessibility– are contextual and depend upon the type of user [3]. User testing is recommended to ensure the accessibility of a Web site. These tests provide quantitative and qualitative data about real users that perform real tasks with a product [4].

In this paper, we focus on web context and propose a new methodology that facilitates the creation of a user test environment (elements to be evaluated, tasks to be performed) adapted to the characteristics of users with disabilities. In this sense, preparing a user test environment aimed at assessing specific barriers for disabled users can be quite an effort for the evaluator preparing the test: selecting barriers affecting each disability to evaluate, choosing specific HTML elements that incorporate the barriers and adequate coding of the elements, deciding how to approach the tasks to be performed, etc.

Without a systematically automated process, there is no way to ensure that the tools used for proper testing incorporate all the elements to guarantee adequate results. Additionally, it hardly ensures an adequate and exhaustive list of tasks to run during the test.

Moreover, accessibility barriers cannot be generalized: accessibility barriers related to mobility-impaired users are different from those associated with visually impaired users, and in each specific case the evaluation environment (Web site/s) should incorporate different elements to ensure the validity of the test.

Hence, accessibility evaluators should raise several questions when designing experimental Web user tests for people with disabilities:

- To which group do my main users belong?
- What elements are especially problematic for this group?
- What tasks should be performed by users on my Web site to assess the impact of problematic elements?
- What Web content should be added to the Web site to allow the user group to carry out tasks?

To address these questions and explain the methodology, the paper is organized as follows. Section 2 describes the related work. The methodology is introduced in Section 3. Section 4 presents a case study and the last section presents the conclusions of the work.

II. RELATED WORK

By barrier, we mean any condition that prevents the execution of an objective by users with disabilities. Initiatives such as the "Accessible Project" [5] and the "Barrier Walkthrough" [6]have studied the most problematic Web barriers affecting different profiles of users with disabilities (visual, low vision, hearing, cognitive and motor). It is important to point out that disabled users interact with the Web in different ways [7] and it is necessary to analyze their characteristics in order to understand which barriers cause a greater impact in each case.

There are methodologies, proposed in [8], [9] and [10] based on the importance of user testing for developing software, but they do not take into account the process of preparing the necessary environment for carrying out the user test. In this sense, the methodology we propose provides a proper framework.

Various methodologies are followed for evaluating the accessibility of a Web site [11], [12]. Traditionally, accessibility evaluation is carried out by applying the Web Content Accessibility Guidelines (WCAG) [13] [14], or its ISO equivalent, ISO/IEC 40500:2012 [15]. However, WCAG guidelines are primarily a legal instrument and, as several authors point out [16] [17] [18], compliance with them does not guarantee 100% accessibility of the Web site. Moreover, according to some authors [19] [20], WCAG 2.0 guidelines do not address all the needs of users for accessing content. Consequently, although WCAG guidelines provide an important starting point, conducting user tests is very important for detecting errors to as great an extent as possible [20] [4].

Given the importance that WCAG guidelines have gained, several tools have been developed to automate the evaluation of Web accessibility [22]. However, not all elements can be evaluated automatically, given that it is necessary to complement the evaluation with a manual review.

In general, automatic accessibility evaluation tools provide result reports that are too technical, making it difficult for people who are untrained in accessibility to understand them. There are some examples that make accessibility evaluation available to the non-expert evaluator, helping that person to understand WCAG guidelines and to create Web content. Accessibility Evaluation Assistance (AEA) [23] [24] is a tool that shows the aspects that must be taken into account for applying WCAG guidelines, thus facilitating the evaluation process. Accessibility Example Generator (AEG) [25] is another tool that assists in the process of creating examples of both accessible and inaccessible Web pages.

Related to all this, the Web Accessibility Initiative (WAI) of the W3C has two projects that work with aspects related to Web content source code: (i) the project WCAG 2.0 Test Samples Repository [26] features accessible and inaccessible code samples within the context of WCAG 2.0 guidelines; and (ii) the project "Before and After Demonstration" [27] is a set of web pages showing examples of accessibility barriers.

In this sense, the methodology we propose is focused on minimizing the effort of creating the user test environment aimed at people lacking a background in accessibility in order to adapt content to the needs of an organization. The tool automates the methodology, optimizes the time and resources needed to create the HTML elements to be tested and streamlines a process which is costly if done manually.

III. METHODOLOGY

The proposed methodology has the following key strengths:

- To optimize the effort for preparing the user test environment: evaluators can reduce the time needed to create code or to analyze the guidelines relevant to their case.
- To reduce the time to prepare and generate the user test environment: evaluators can minimize the creation time of the test environment by using a tool that automates the methodology.
- To ensure the inclusion of all the HTML elements to be tested from a selection of profiles and HTML elements, specific accessibility barriers are listed for assessment, as well as tasks likely to cause access problems for previously selected user profiles.
- Ease of use: accessibility related information is presented in a simple and understandable manner for users without prior knowledge of accessibility.

Below (Figure 1) is an outline of the methodology presented.

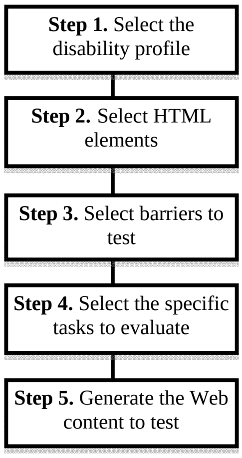


Figure 1: Outline of the methodology presented.

A. *Description of the methodology*

The methodology uses and lists several data sources: disabilities, Web page elements, accessibility barriers and WCAG guidelines, which are put together to obtain a list of tasks to run on a web page with accessible and inaccessible source code. The methodology consists of five steps in which the evaluator selects the user profile of the test (step 1), HTML elements that present problems to the selected user (step 2), barriers to testing (step 3) and specific tasks to perform in the test (step 4) to finally obtain a source code with a web implementation that collects all aspects of the previous selections (step 5). The evaluator has the code that best fits the needs of the user test or that best performs the test using all the information resources generated.

1) Step 1: Select the disability profile: In the first step, the characteristics of the user who will participate in the test are selected. The user profiles offered by the International Classification of Functioning, Disability and Health (ICF) [28] were considered, but we thought that it was more appropriate to group the characteristics into a total of four groups further split by the severity of disability.

1. People with mobility impairments: 1a) mild and moderate limitation and 1b) severe limitation.
2. People with cognitive impairments: 2a) mild and moderate limitation and 2b) severe limitation.
3. People with hearing impairments: 3a) mild and moderate limitation and 3b) severe limitation or total deafness.
4. People with visual impairments: 4a) low vision, 4b) severe limitation or blindness and 4c) color-blind.

2) Step 2: Select HTML elements: All HTML elements that are of interest to test and that cause some problem for the users of the profiles chosen in step 1, are selected.

3) Step 3: Select barriers to test: The accessibility barriers to be tested are selected in the next step. Only the barriers that cause problems are listed according to the HTML elements and the user profiles previously chosen. We have used the following information sources to obtain the list of barriers: the accessible project [5], which lists the disability accessibility barriers that most directly have an impact and the work of Cunningham [29], which lists the most problematic HTML elements for each type of disability.

Keep in mind that the same barrier can affect one or more of the various disabilities.

4) Step 4: Select the tasks to evaluate: The possible tasks to perform in the user tests according to previous selections are chosen in this step of the methodology. As a barrier can affect one or more disabilities, the principle of the task is different for each affected group. For example, the barrier "moving content" is perceived differently depending on whether the user has visual or cognitive disabilities. Therefore, an adequate approach for participants with disabilities is essential.

5) Step 5: Generate the Web content to test: The last step of the methodology is to obtain the Web content (accessible and inaccessible), which will be incorporated into the pages to test, in keeping with the previously selected parameters.

This may not be the final code since its aim is to inspire and facilitate the task of creating a test page by the team responsible for carrying out the user test.

The methodology also offers information on how to evaluate the accessibility barriers in order to supplement the information given to a new evaluator.

IV. CASE STUDY

In the following section, we show a case study that implements the previously proposed methodology through a mock-up of the tool. This case study was based on an evaluation by the authors of the paper [31], rather than with actual evaluators with no technical knowledge. The examples of web pages were inspired by W3C WAI and provided by [32].

It begins with the premise of running a user test for people with severe visual impairments or blindness and to find out the difficulties that a contact form can cause.

A. *Steps of methodology*

Below are the steps to be followed:

1) Step 1: Select the disability profile: A user test is prepared, adapting it to a specific user profile. As an example, the following profile is selected: People with severe visual impairments or blindness. To empathize with these users and to better understand the difficulties they encounter, a profile of a user with the selected disability is displayed.

TABLE I. HTML ELEMENTS THAT AFFECT USERS WITH VISUAL DISABILITIES AND THE WCAG1 AND WCAG2 REGULATIONS RELATED TO THEM.

HTML element	WCAG 1.0	WCAG 2.0
Browsing		
Shortcuts	13,6	2.4.1
Keyboard shortcuts	9.5	--
Design		
Color	2.1	1.4.1
Header/page title	13.2	2.4.2
Tables as layouts	3.3, 5.3, 5.4	1.3.1, 1.3.2
Structure		
Headings (H1, H2)	3.5	2.4.10, 1.3.1,
Frames	12.1 12.2	2.4.1, 4.1.2
Popups	10.1, 9.4, 9.5	3.2.1, 3.2.5
Language definition	4.1, 4.3	3.1.1, 3.1.2
Content		
Images	1.1, 3.1	1.1.1, 1.4.5, 1.4.9, 1.3.1
Opaque objects	1.1	1.1.1
Moving or blinking content	1.1, 7.2, 7.3	1.1.1, 2.2.2
Links	10.5, 12.3, 13.1	2.4.4, 2.4.9, 2.4.10
Data tables	5.1, 5.2, 5.5, 5.6, 10.3	1.3.1
Multimedia: Video	1.1, 1.4	1.2.1, 1.2.2, 1.2.3, 1.2.4, 1.2.5, 1.2.7, 1.2.8, 1.2.9
Forms	5.3, 7.5, 9.4, 10.2, 10.3 12.4, 12.3	1.3.1, 1.3.2, 2.4.3, 2.4.7, 4.1.2, 2.4.10
Text	3.1	1.4.5, 1.4.9
Time limit	7.4, 7.5	2.2.1, 2.2.4, 3.2.5
Javascript/Flash	6.3, 6.4, 6.5, 8.1	2.1.1, 2.1.3, 4.1.2
Events		
Dynamic Changes	6.3 8.1	4.1.2 3.3.1, 3.3.2, 3.3.3
Event controllers Mouse events and keyboard	6.3, 6.4, 9.3	2.1.1, 2.1.2, 2.1.3

2) Step 2: Select the HTML elements: The elements on which the test is focussed are selected.TABLE I. shows a list of all problematic elements that affect users with severe visual impairment. Each element is accompanied by related WCAG guidelines. This information is only needed in order to be able to list different information groups internally, and no information would be displayed in the user interface. In our case we selected the item "Form" to test the particular difficulties users may have in sending a contact form. To facilitate the communication of the Web element, each element is accompanied by an HTML example that may be incorporated into the test (See Figure 2).

```
<form id="form1" name="form1" method="post"
action="">
<legend>Registration</legend>
 <div>
  <span> <label for="txt”>Name <label> </span>
  <span> <input type="text" name="txt" id="txt"/>
  </span>
 </div>
</form>
```

Figure 2. A form example.

3) Step 3: Select barriers to test: With the previously selected information the accessibility barriers to be tested are shown. As a specific example, TABLE II. shows the barriers related to visually impaired users (step 1) and the form elements (step 2).

TABLE II. LIST OF BARRIERS RELATED TO VISUALLY IMPAIRED USERS AND THE WEB "FORM" ELEMENTS

Barriers	Description
Forms that direct the user to a different page	The form is updated and loses the previously entered information when selecting the send button.
Forms with fields not semantically marked	Control labels (LABELs) must describe the data to be entered in each textbox.
Forms that are not correctly displayed on a screen reader	The elements of the label form and textbox are organized properly. When the user browses with the tabs, the user will find the following fields: Name (TAB) name textbox. (correct)

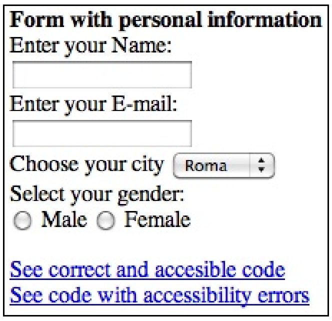
4) Step 4: Select the tasks to evaluate: A list of tasks permits the evaluation of previously selected barriers and elements, adjusted to the specified user profiles.TABLE II. shows only the information concerning the barrier "Forms with fields not semantically marked".

5) Step 5: Generate the Web content to test: A proper Web code is displayed for each selected Web element. TABLE IV. shows an example of the barrier "Forms with fields not semantically marked". Example web pages inspired in W3C WAI activity are provided by [32].

TABLE III. LIST OF TASKS CONCERNING SEVERE VISUALLY IMPAIRED USERS AND THE WEB FORM ELEMENT AND THE BARRIER "FORMS WITH FIELDS NOT SEMANTICALLY MARKED".

Barrier	User task
Forms with fields not semantically marked	Fill in a form with personal information (name, email, city and gender)

TABLE IV. WEB CONTENT ASSOCIATED WITH THE BARRIER "FORMS WITH FIELDS NOT SEMANTICALLY MARKED" AND THE TASK "FILL IN A FORM WITH PERSONAL INFORMATION."

Content description
Each form field must be labeled with a label element. Additionally, it must be identified with a unique name that is also included in its corresponding label element.
Content example

Verification of the barrier
Manual verification: To create the form we suggest the use of a form-builder tool. To validate the form code we suggest using the WAT tool: it displays the text fields that are identified with labels and IDs in the same form.

B. *Create the user test environment*

Once the HTML elements are known, as well as the specific Web content to be added to the website, the next step is to proceed with its creation and implementation. At the end of the process, a summary is shown (See TABLE V.). The evaluator can embed an obtained code in a simple web page and start the test.

The authors of this article have checked the validity of the proposed methodology by creating several user tests aimed at people with cognitive disabilities [30] and visual disabilities [31]. By following the methodology, there is a reduction in the time needed for creating the user test environment and an improvement in the effectiveness of preparing the test environment to include the most problematic elements for users with disabilities. The studies carried out by the authors were focused on learning the mental state of users with disabilities as they interacted with each barrier to accessibility. It was necessary to be very precise when creating the HTML files to be evaluated.

TABLE V. SUMMARY OF STEPS TO CREATE USER TEST ENVIRONMENT

Summary of steps

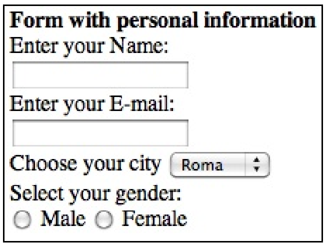
If you want help in creating the web page, feel free to copy some of the example pages provided.

Your settings:

- You selected blind and severe **visual disabilities group**
- You focused on **form elements**
- You selected the barrier "**Forms with fields not semantically marked**"
- A suggested task is to **ask users to fill in a form** with personal information (name, email, city, gender)

You could test the impact of this accessibility barrier with this user profile comparing a non-accessible implementation with an accessible implementation, both with an exact visual appearance, as follows.

Visual appearance



Source code (Accessible)

```
<legend>Registration</legend>
<form id="form1" name="form1" method="post"
action="">
 <label for="nom">Enter your Name:</label>
 <input type="text" name="txtnom" id="nom" />
 <label for="mail">Enter your E-mail:</label>
 <input type="text" name="mail" id="mail" />
 <label for="favcity">Choose your city</label>
  <select id="city" name="select">
   <option value="1">Roma</option>
   <option value="2">Paris</option>
   <option value="3">London</option>
  </select>
 <legend>Select your gender</legend>
 <input id="Male" type="radio" name="sen"
    value="Male">
  <label for="Male">Male</label>
  <input id="Female" type="radio" name="sen"
    value="Female">
  <label for="Female">Female</label>
</form>
```

Source code (Non Accessible)

```
Registration<br/>
<form id="form1" name="form1" method="post"
action="">
 <label>Name:</label>  <input type="text" />
  <label>E-mail:</label> <input type="text" />
 <span><label for="favcity">City</label>
 <label>
   <select name="select2" id="select">
     <option value="1">Roma</option>
     <option value="2">Paris</option>
     <option value="3">London</option>
   </select>
 </label>
 <label>Gender:</label>
 <input type="radio" name="radio1"
   id="radio1" /> Male
 <input type="radio" name="radio1"
   id="radio1" /> Female
</form>
```

V. CONCLUSIONS

This paper presents a new methodology that optimizes the design of a user test environment for people with disabilities. This methodology is shown through a case study. The proposed methodology is based on following a series of specific steps in order to provide suitable content and tasks for carrying out a user test for people with disabilities. The evaluators can conduct the user test without specific knowledge of accessibility, the methodology minimizes the effort needed for creating such an environment adapted to the specific needs of the organization. The methodology is accompanied by a tool that automatically performs some of the steps and speeds up implementation, thus reducing time and costs in creating the right environment for user tests. In the case study, with only five selections (and five mouse clicks) the evaluator has a suggestion of task, and the code of both accessible and non-accessible implementations.

The authors followed this methodology with users with cognitive disabilities [30]. As a conclusion we found that using the proposed methodology facilitates the inclusion of elements to be tested and optimizes the time and available resources used in the preparation of the test.

In a complementary manner, the methodology can also help webmasters to be aware of the most problematic elements (barriers) of a specific disability and with this knowledge, they are more aware of accessibility requirements in a Web page.

ACKNOWLEDGEMENTS

This work was financed by the project, Improved User Experience for Sustainable Open Government Data Infrastructures (InDAGuS-UX) (TIN2012-37826-C02-02) of the Ministry of Economy and Competitiveness and by the collaboration agreement signed between ACTIVA MÚTUA 2008 and the University of Lleida to develop technologies and projects that enable labour adaptation and reintegration of people with disabilities in the field of ICT. Funding has also been provided by the pre-doctoral grant from the University of Lleida awarded to Afra Pascual.

REFERENCES

[1] World Health Organization (WHO). "World report on disability" (2011). Retrieved January 2014, from http://www.who.int/disabilities/world_report/2011/report/en/index.html

[2] M. Hertzum. "Problem prioritization in usability evaluation: from severity assessments toward impact on design". International Journal of HumanComputer Interaction 21, 2 (2006), pp. 125-146.

[3] A. Donelly and M. "Maggenis. Making Accessibility Guidelines Usable", LNCS 2615, 2003. pp. 56–67

[4] S.L. Henry. "Just Ask: Integrating Accessibility Throughout Design". Retrieved January 2014, from www.uiAccess.com/JustAsk/

[5] Accessible: Accessibility Assessment Simulation Environment for New Applications Design and Development Accessible.(ICT programme of FP7). Retrieved January 2014 from: http://www.accessible-eu.org/

[6] G. Brajnik. “Web accessibility testing with barriers walkthrough”. Retrieved January 2014 from: www.dimi.uniud.it/giorgio/projects/bw

[7] S. Abou-Zahra. “How People with Disabilities Use the Web”. ed. Copyright © 2012 W3C® (MIT, ERCIM, Keio). Retrieved January 2014 from: http://www.w3.org/WAI/intro/people-use-web/Overview

[8] J. Nielsen. “Usability Engineering”. Academic Press Professional, Boston, MA, USA, 1993

[9] J. Rubin. “Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests”. John Wiley& Sons, NY. 1994

[10] H. Petrie, F. Hamilton. N. King, and P. Pavan. “Remote usability evaluations With disabled people”. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06), Rebecca Grinter, Thomas Rodden, Paul Aoki, Ed Cutrell, Robin Jeffries, and Gary Olson (Eds.). ACM, New York, NY, USA, pp.1133-1141.

[11] E. Velleman and S. Abou-Zahra. “Website Accessibility Conformance Evaluation Methodology” (WCAG-EM) 1.0. Retrieved January 2014 from http://www.w3.org/TR/WCAG-EM/

[12] E. Velleman, C. A. Velasco, M. Snaprud, and D. Burger. “D-WAB4 Unified Web Evaluation Methodology” (UWEM 1.0). Technical report, WAB Cluster, 2006.

[13] W. Chisholm, G. Vanderheiden, and Y. Jacobs. “Web content accessibility guidelines 1.0”. Retrieved January 2014 form: http://www.w3.org/TR/WCAG10/

[14] B. Caldwell, M. Cooper, L. Reid, and G. Vanderheiden. “Web content accessibility guidelines 2.0”. Retrieved January 2014 from: http://www.w3.org/TR/WCAG20/

[15] ISO/IEC 40500:2012. “Information technology -- W3C Web Content Accessibility Guidelines (WCAG) 2.0” (2012) Retrieved January 2014 from: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=58625

[16] D. Rømen and D. Svanæs. “Evaluating web site accessibility: validating the WAI guidelines through usability testing with disabled users”. In Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges (NordiCHI '08). ACM, New York, NY, USA, pp. 535-538.

[17] C. Harrison and H. Petrie. “Severity of usability and accessibility problems in eCommerce and eGovernment websites”. In Computers and People XX: Proceedings of British Computer Society Human Computer Interaction Conference (BCS HCI’ 06), pp.255- 262. London: British Computer Society.

[18] M. Ribera.; M. Términs, and M. García-Martín. “How to perform usability tests with blind people”. El professional de la información.Vol.17, nº 1 / January - February 2008. pp. 99 - 105

[19] C. Power, A. Freire, H. Petrie, and D. Swallow. “Guidelines are only half of the story: accessibility problems encountered by blind users on the web”. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12). ACM, New York, NY, USA, pp. 433-442. DOI=10.1145/2207676.2207736

[20] S. Leuthold, J. Bargas-Avila, and K. Opwis. “Beyond web content accessibility guidelines: Design of enhanced text user interfaces for blind internet users”. Int. J. Hum.-Comput. Stud. 66, 4 (April 2008), pp. 257-270. DOI=10.1016/j.ijhcs.2007.10.006

[21] “BSI. Guide to good practice in commissioning accessible websites”, BSI PAS 78 (2006). Retrieved January 2014 from: http://www.bsi-global.com/en/Shop/PublicationDetail/?pid=000000000030129227

[22] S. Abou-Zahra. “Complete List of Web Accessibility Evaluation Tools”. Retrieved January 2014 from: http://www.w3.org/WAI/ER/tools/complete

[23] C. Bailey and E. Pearson. “An educational tool to support the accessibility evaluation process”. In Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A) (W4A '10). ACM, New York, NY, USA. DOI=10.1145/1805986.1806003

[24] C. Bailey and E. Pearson. “Development and trial of n educational tool to support the accessibility evaluation process”. In Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (W4A '11). ACM, New York, NY, USA, , Article 2 , 10 pages. DOI=10.1145/1969289.1969293

[25] A. S. Al-Khalifa and H.S Al-Khalifa. “An educational tool for generating inaccessible page examples based on WCAG 2.0 failures”. In Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (W4A'11). ACM, New York, NY, USA. Article 30, 4 pages. DOI=10.1145/1969289.1969328.

[26] S. Abou-zahra and M. Cooper. “WCAG 2.0 Test Samples Repository”. In Proceedings of the 5th International Conference on Universal Access in Human-Computer Interaction. Part III: Applications and Services (UAHCI '09), Constantine Stephanidis (Ed.). Springer-Verlag, Berlin, Heidelberg, pp. 619-627. DOI=10.1007/978-3-642-02713-0_65

[27] BAD – “Before and After Demonstration”. (2005) Retrieved January 2014 from: The World Wide Web Consortium (W3C). http://www.w3.org/WAI/EO/2005/Demo/

[28] International Classification of Functioning, Disability and Health ICF (2013). Retrieved January 2014 from: http://www.who.int/classifications/icf/en/

[29] K. Cunningham. “The Accessibility Handbook”. (2012) ISBN: 978-1-449-32285-4

[30] A. Pascual, M. Ribera, and T. Granollers. “Degree of involvement of accessibility barriers web in users with intellectual disabilities”. Actas del XIV Congreso Internacional de Interacción Persona-Ordenador (INTERACCIÓN 2013) ISBN: 978-84-695-8352-4

[31] A. Pascual, M. Ribera, and T. Granollers. “Impact of accessibility barriers on the mood of blind, low-vision and sighted users”. Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. (ISSN: 1877-0509) (2013)

[32] S. About-Zahra. Before and After Demonstration (BAD) http://www.w3.org/WAI/demos/bad/after/survey.html

Comparing Recognition Methods to Identify Different Types of Grasps for Hand Rehabilitation

Beatriz Leon, Angelo Basteris and Farshid Amirabdollahian
Adaptive Systems Research Group
University of Hertfordshire
Hatfield, Hertfordshire, UK
Email: b.leon, a.basters, f.amirabdollahian2@herts.ac.uk

***Abstract*—Grasping activities are extremely frequent in the set of activities of daily living. This causes severe impairments for stroke survivors, whose wrist and hand may suffer from a variety of symptoms such as spasticity, hypertone and muscular weakness. Intensive repeated movement performance is at the base of robot-therapy. Thus, patients may benefit, in terms of functional recovery, from the integration of grasp gestures in robot mediated exergaming. In this feasibility study, we developed and tested three methods for recognizing four different grasp postures performed while wearing an exoskeleton for hand and wrist rehabilitation after stroke. The three methods were based on the statistics of the produced postures, on neural networks and on support vector machines. The experiment was conducted with healthy subjects, with no previous injuries on the hand, during grasping of actual objects and then repeated using imaginary objects. We compared the three methods in terms of accuracy, robustness with respect to the size of the training sample, inter-subjects' variability, differences between different postures and evaluating the presence of real objects. Our results show that the support vector machine method is preferable in terms of both accuracy and robustness, even with a small training sample, with training times on the order of milliseconds.**

***Keywords*—grasp posture recognition; stroke rehabilitation; Support Vector Machines; Neural Networks.**

I. Introduction

Stroke survivors are often unable to perform tasks requiring fine motor control, which are needed during activities of daily living, among which grasping is one of the most recurrent. Robots represent an excellent tool for exercise-based approach in neurorehabilitation, but in order to increase the functional outcome of the treatment patients training should consist in the performance of grasping movements similar to those needed in daily life [1].

The Supervised Care & Rehabilitation Involving Personal Tele-robotics (SCRIPT) project aims at delivering an affordable system for home-based rehabilitation of the hand and wrist for stroke survivors [2]. An exoskeleton has been developed within the project in order to facilitate the patients' fingers and wrist extension, due to observed abnormal flexion synergies often leading to flexed hand and wrist [3]. The motivation of the participant is enhanced by therapeutic exercises mediated via interactive games, which subjects control by wearing the orthosis and performing several movements of the whole upper limb, from the shoulder to the fingers.

There are numerous types of grasping. Feix et al. refer to 17 gross types of grasping [4]. These are utilised within the daily life interaction. Our aim is to incorporate the detection of these grasps in the rehabilitation framework so that they can be incorporated into human-machine interaction with the aim of increasing interaction time.

The problem of hand posture detection has been approached with several methods. A preliminary gross distinction can be made between vision-based or glove-based approaches. Vision based approaches allow natural hand gesture recognition, typically by detection of the fingertips. Detailed information about such type of technologies can be found on comprehensive reviews such as Chaudary et al. [5]. Among the vision-based techniques, a possibility is to use specific color patterns on a fabric glove in order to facilitate the detection [6]. Glove-based approaches reduce the need for computational power at the cost of possibly altering the natural performance of the movements by making one wear an instrumented glove. This could not be a concern when the presence of such a device is however required in order to assist the patient in movement performance. Several methods have been proposed for hand posture detection using data gloves [5], [7], including feature extraction (e.g., principal component analysis, linear discriminant analysis) and learning algorithms (e.g., neural networks, support vector machines). Specifically, Xu et al. [8] allowed hand gesture recognition for driving by training and testing a neural network with a pattern of 300 hand gestures. Other studies have also compared different approaches for grasp detection. Palm et al. [9] compared three methods: difference between the grasp and models grasps, qualitative fuzzy models and Hidden Markov Models (HMM), concluding that the first method outperforms the other two in terms of accuracy on a set of 10 grasps for 15 different grasps primitives. All the aforementioned studies used a data glove, which measures the angles of 18 joints in the hand (CyberGlove [10]).

In the context of this study, the vision-based approaches are unsuitable given that the exoskeleton causes the visual occlusion of most of the hand. Therefore, the gesture recognition should rely on the sensor readings provided by the device. Additionally, a basic requirement for facilitating the application of grasp recognition in the rehabilitation framework, possibly affecting the future use of the system, is a short setup time.

Another aspect relates to the involvement of real or imag-

 ISBN: 978-1-61208-325-4

inary objects while performing the grasp postures. Ideally, patients should mimic the graping posture without interaction with real objects while playing the games. However, there could be substantial differences between the hand posture observed when holding actual objects and its performance based on motor imagery, which can affect the therapy outcome. On the one hand, the use of real objects would possibly enhance the skills transfer from the training activity to the functional use. Also, grasping real objects would induce physical constraints on the hand posture, which might facilitate the gesture recognition, making its performance more repeatable. However, having real objects also carries disadvantages such as introducing additional requirements in the device design, having the participants focusing their gaze and attention on the objects rather than on the screen and reducing the usability of the system.

Hence, in this work, we test the trade-off between accuracy and training sample size of three different approaches: recognition based on statistics, based on neural networks and based on support vector machines, for three type of grasps, either using actual objects or imagined objects while performing the gestures.

The remainder of the paper is structured as follows. We start with the methods section introducing the passive orthosis used in the SCRIPT project, the selected grasp postures, the different methods selected to recognize the postures and details of the experimental protocol. Then, we report and analyze the results, and conclude with a brief summary of our findings, including the best method found for grasp posture recognition and directions for future work.

II. Methods

A. Measuring device

The SCRIPT passive orthosis [11] is a passive device which features five leaf springs, which are connected to finger caps through an elastic cord (Figure 1). The extension forces resulting from the parallel of these two elements are applied at the fingertip. The elastic cord enables the fingers freedom of movement relative to the leaf spring and also allows to adjust the level of support provided by the device. The leaf springs are fitted with commercial resistive flex sensors [12], which measure their deflexion with a resolution of 1 degree. Because of the elastic coupling, the deflection of the leaf spring is not the actual flexion angle of the finger. However, the two quantities are related by a monotonically increasing function [11]. Also, movements of lateral abduction/adduction of the fingers and opposition of the thumb are not restricted nor sensed by the orthosis. It measures only overall finger flexion in a range from 0 to 90 degrees and wrist flexion and extension in a range from 90 to -45 degrees.

B. Hand Gestures

We selected three types of grasps shown in Figure 2. Two are classified as precision grips: the three-jawed chuck (the thumb opposes the index and middle finger) and the lateral prehension (the thumb holds an object against the side of

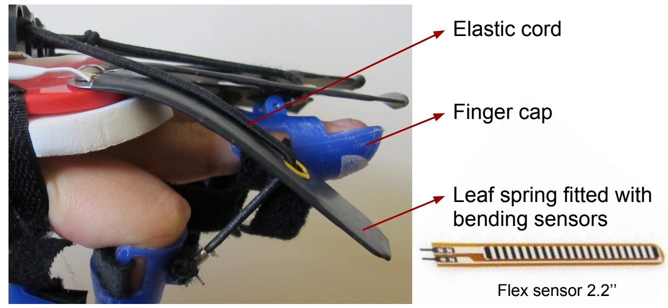


Fig. 1. Bending sensors and the leaf springs of the SCRIPT passive orthosis

the index finger). The third grasp selected is classified as a power grasp: the cylindrical prehension in which all fingers make contact with the object. Keller et al. [13] identified the three-jawed chuck and the lateral prehension as the most frequently used prehensile patterns for static and dynamic grasping, respectively. Additionally, the three-jawed chuck and the cylindrical prehension are tested as part of the Grasp and Grip sections of the Action Research Arm Test [14], which has been used as a reliable, valid measure of arm motor status after stroke. The relaxed posture of the hand was used as the forth gesture in order to be able to recognize when the patients are not performing any grasps. Furthermore, these gestures were selected considering that patients with different levels of hand impairment should be able to perform them.

C. Methods for Recognition

1) Recognition based on statistics of the training samples: The method considers the absolute error of each finger with respect to the average value of the training samples. In the training phase, the mean value of the flexion measured for each finger f is calculated for the N number of training samples:

$$mean_f = \frac{\sum_{i=1}^{N} |Flexion_f i|}{N} \tag{1}$$

During the testing phase, the flexion values of each finger are compared with the mean value of the training phase and averaging it among the five fingers, identifying then the gesture only if this value falls below a threshold th:

$$\frac{\sum_{f=1}^{5} |Flexion_f - mean_f|}{5} \leq th \tag{2}$$

The value of th was empirically set to 10 degrees as this value allowed to reach an accuracy of 90% when tested on a single subject during a pilot study.

With this method, gestures might be recognized even though one or more fingers are in very different conditions from what measured in the training phase, provided that other fingers compensate for this overall difference.

2) Recognition based on Neural Networks (NN): Artificial Neural Networks (NN) [15] have been extensively used for supervised learning to solve problems of pattern recognition (classification) and regression [16]. They have been previously used for hand posture classification [5], [17].

(a) Three-jawed chuck (b) Lateral prehension (c) Cylindrical prehension (d) Relaxed posture

Fig. 2. Selected gestures to be recognized. Bottom images show how they are performed wearing the SCRIPT device.

In this work, we used a three layer neural network that contain five input nodes for the flexion values of each finger, 10 hidden nodes and four output nodes, one for each gesture to be recognized (the three selected gestures plus one relaxed postured acquired by instructing the subjects to relax their fingers). The flexion angles were normalized in a range from 0 to 1, corresponding to 0 to 90 degrees. Similarly, the limits of the output nodes were set to 0 and 1. We used back propagation, a learning mechanism that minimizes the error between target output and the output produced by the network until it reached an error of 0.01.

After the training phase, a gesture was recognized if the results given by an output node was higher than 0.7 and the other gestures has a recognition rate less than 0.3. Otherwise, no gesture was returned by the model for the given posture. This method was implemented in Python using the *Neurolab* library [18].

3) Recognition based on Support Vector Machines: This method utilises Support Vector Machines (SVM), which is a popular machine learning technique for classification [19]. A support vector machine constructs a set of hyperplanes in a high-dimensional space that are used to classify the data. A good separation is archived by the hyperplane that has the largest distance to the nearest training data point of any class. The hyperplanes are found solving the following optimization problem [20]:

$$\min_{w,b,\xi} \quad \frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i \tag{3}$$
$$\textit{subject to} \quad y_i(w^T\phi(x_i) + b \geq 1 - \xi_i) \textit{ and } \xi_i \geq 0$$

where $\{(x_i, y_i)|x_i \in R^p, y_i \in -1, 1\}$ are the training set of l instance-label pairs, x_i is p-dimensional real vector, w the normal vector of the hyperplane and $C > 0$ the penalty parameter of the error term. The training vectors x_i are mapped into a p-dimensional spaces by the function ϕ and in order to create nonlinear classifiers a kernel function is used. In our work, we used a radial basis function (RBF) as the kernel function, given that it can handle the case when the relation between class labels and attributes is non-linear. It is defined as:

$$K(x_i, x_j) = exp(-\gamma \parallel x_i - x_j \parallel^2), \gamma > 0 \tag{4}$$

where γ is a kernel parameter.

Therefore, two parameters are needed: C and γ. In order to find the best values for this parameters, we used a $v - fold$ cross-validation technique, dividing the training set for one subject into v subsets of equal size and testing the classifier on the remaining $v-1$ subsets. The values found were: $C = 4$ and $\gamma = 1$. The method was implemented in Python using the *LIBSVM* package [20]. As with the previous method, the flexion angles were normalized in the range from 0 to 1. The selected error to stop the training face was 0.01.

D. Experimental protocol

This work was designed as a feasibility study aimed at comparing different methods for grasp posture recognition and selecting one that can be further adopted in rehabilitation. We decided to perform the experiments of this study on healthy subjects, while focusing on the recognition capabilities of the different methods.

Five healthy subjects (age = 31 ± 2.1, 3 males/2 females) with no previous injuries of fingers, hand or wrist volunteered to participate in this study. Participants were recruited amongst faculty staff by advertising on an internal mailing list. All subjects were right handed.

This study was carried out at the University of Hertfordshire and approved by the university ethics committee (Ethics protocol number COM/ST/UH/00008).

The participants were asked to wear a medium sized left hand SCRIPT passive orthosis while sitting in front of a PC. Participants were instructed to grasp one out of three objects (a ball, a mug and a notepad) by showing on the screen the picture of the appropriate grasp (Figure 2). The subject then confirmed with a click that he/she achieved the desired gesture and the flexion angles of the fingers were saved. After confirmation, they were asked to release the object, relax the hand and press a button. At that moment, the flexion angles of the fingers of the relaxed posture were also saved.

Each subject performed six repetitions of each gesture in a pseudo-random sequence grasping the real objects. Subsequently, participants repeated the same procedure but mimicking the required gesture without actual objects (the difference is shown in Figure 3 for the first grasp).

Data were then post-processed by Python ad-hoc applications implementing the three methods. Each technique was evaluated based on its computational time and accuracy of recognition, defined as the total number of correctly classified gestures over the total number of gestures.

The data acquired for each subject was divided into two sets for training and testing purposes. The number of training samples (N) was varied between 2 and 5 samples per gesture

 ISBN: 978-1-61208-325-4

(a) (b)

Fig. 3. Example of performing a gesture while: a) grasping a real object or b) grasping an imaginary object

type. The training set was then used to train the model according to each method, which allows us to provide insights into what is the minimum number of samples required to achieve a given accuracy. Results were considered taking into account all possible permutations of the training samples, for each training sample size N. Data analysis was done using IBM SPSS for Windows version 21.0.

III. Results

A. Recognition methods

Figure 4 shows the results in terms of accuracy for the three methods, averaging all subjects. Regardless of the training sample size, the SVM approach outperforms the other methods in terms of accuracy, with median values greater than 90% already with two repetitions of a gesture. While the method based on statistics did not allow achieving comparable accuracies, neural networks can potentially be used. However, this would happen at the cost of increasing the training sample.

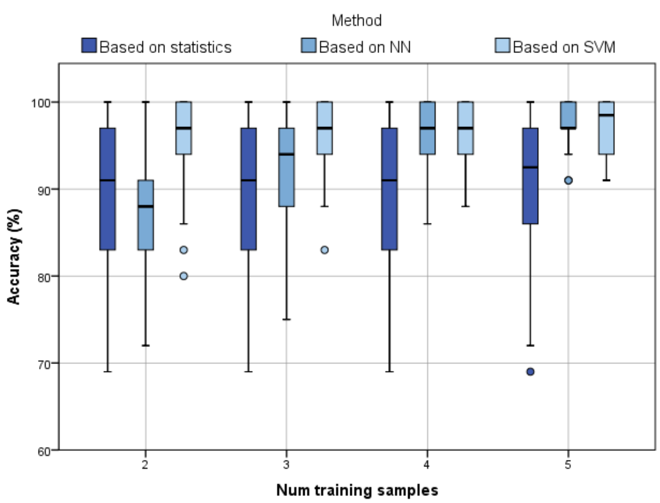


Fig. 4. Accuracy for each method tested over all subjects.

The computational times taken for each method is presented in Table I. The method based on statistics took less than 1 ms regardless of the training sample size, given that it involves only simple arithmetical and logical operations. Similarly, the method based on SVM needed a small training time (below 56 ms for all conditions). On the contrary, the method based on NN required much higher durations, ranging from 0.1 s to several hours.

TABLE I
TIME REQUIRED TO TRAIN THE DIFFERENT METHODS

	Computational time (sec)		
Method	Mean	Maximum	Minimum
Based on statistics	<0.001	<0.001	<0.001
Based on NN	89.543	>10000	0.110
Based on SVM	0.005	0.056	0.001

Therefore, although the method based on NN was able to reach median recognition accuracies higher than 90% with more than 3 training samples, the long training time, and particularly its high variability, suggested to drop it in favour of the method based on SVM, which was used to perform the further analyses evaluating the presence of real objects and the variability among subjects.

B. Grasping real vs. imaginary objects

The results of comparing the accuracy of gesture recognition using real or imaginary objects are shown in Figure 5 for the method based on SVM. The gestures performed without objects are systematically less recognized than the ones using objects.

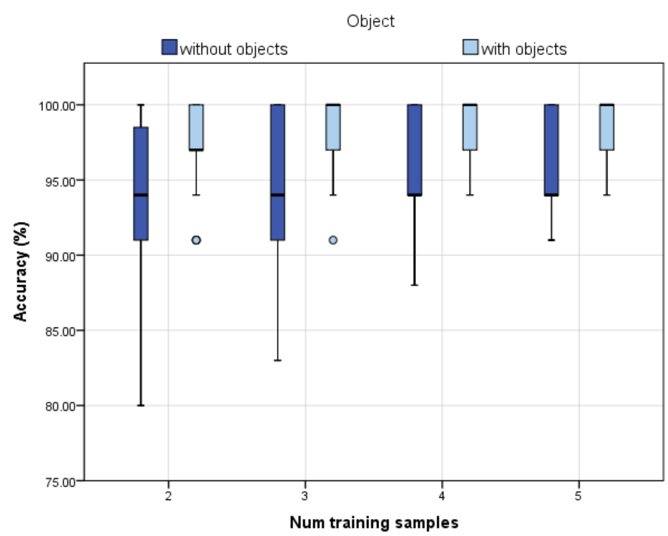


Fig. 5. Comparison of the accuracy of gesture recognition when grasping objects or imagining them, using the method based on SVMs.

This result was expected as the gestures grasping a real object are performed in a more consistent manner than when the objects are imagined. However, performing the gestures using real objects while playing a game is not practical, therefore the use of imaginary objects is advisable as the median accuracies of recognition are no lower than 90%.

C. Number of training samples required

It is important to determine the number of required training samples needed to achieve a good accuracy of recognition.

 ISBN: 978-1-61208-325-4

As the patients need to perform the training procedure before starting the games, the less number of samples required to train the model the better.

For the method based on SVMs, Figure 5 shows the results of the accuracy for different number of training samples acquired per gesture. The results show that the accuracy of recognition increases with the number of training data, as expected. We performed a contrast test between the different number of samples used for training (Table II). The results show that for grasping both real and imaginary objects the means of the groups are significantly different (<0.05) and that there is no significant difference increasing from 4 to 5 samples ($p > 0.05$), therefore we recommend using 4 samples per each gesture.

TABLE II
CONTRAST TEST OF THE EFFECT OF THE NUMBER OF TRAINING DATA OVER THE GESTURE RECOGNITION ACCURACY

Object	Means of groups significance	Contrast significance		
		2 vs. 5 samples	3 vs. 5 samples	4 vs. 5 samples
With	0.007	0.004	0.048	**0.375**
Without	<0.001	0.000	0.004	**0.174**

D. Variation of recognition per gesture

In this section, we considered the variation of accuracy among gestures. The results of the accuracy using the method based on SVMs and 4 samples for training are shown in Figure 6. It can be seen that the relaxed posture is always correctly recognized. The three-jawed chuck has also very high accuracy, despite the presence of a few outliers. The cylindrical and lateral prehension appear as those more difficult to be recognized, specially when using imaginary objects. This is likely because of their similarity, particularly enhanced by the constraints on movements imposed by the orthosis.

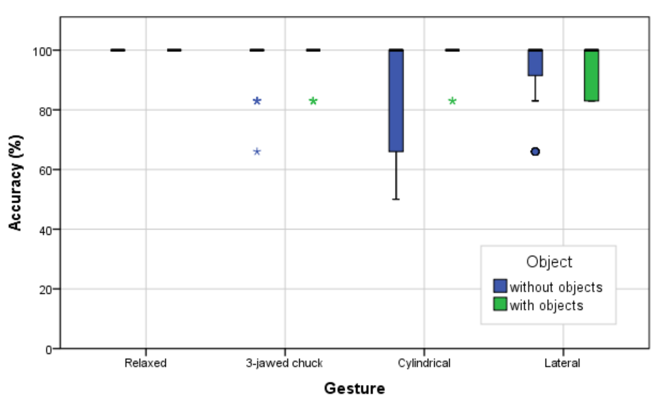


Fig. 6. Comparison of the recognition accuracy obtained for each of the selected gestures using the method based on SVMs.

E. Variation between subjects

In order to study the variation between subjects, we have first analyzed the differences in the flexion angles for each of the gestures performed by the 5 subjects while imagining the objects and then we analyzed how this variation on the flexion angles inter subjects affected the gesture recognition.

The finger flexion angles for each subject varied in most of the cases less that 12 degrees, except for the small finger that presents variation up to 21 degrees performing the lateral prehension for subject 3 and the thumb for subject 5 with a standard deviation of 19 degrees. Figure 7 shows a summary of these results over all subjects per grasp gesture. As it can be expected, the relaxed posture is the most consistent over all the gestures. For the 3-jawed chuck, the ring and small fingers present high variation (> 60 degrees) as they were not directly involved in the gesture as well as the thumb, which flexion reading can vary according to the abduction position which is not measured by the device. The cylindrical prehension shows approximately the same variation over all fingers and the lateral prehension show the highest variation since the flexion of the fingers is not highly constrained by the grasp, as long as all the fingers are flexed and the thumb is above the index finger. These results are correlated with the recognition accuracies of each gesture presented in Figure 6, which shows that the cylindrical and lateral prehension could have low recognition accuracies (up to 50%) given their high variation inter-subjects.

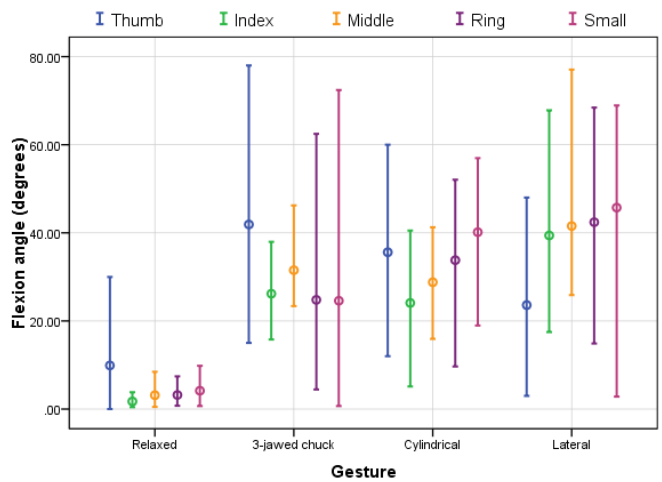


Fig. 7. Variability of the different finger flexion produced by all subjects while performing the different gestures

Finally, Figure 8 presents the accuracy of recognition for each subject (using SVM, without objects). The results shows small variation between subjects and an accuracy higher than 90% for all of them using 3 or more training data per gesture.

IV. CONCLUSION AND FUTURE WORK

In this paper, we evaluated three different methods to classify hand gestures while wearing a passive orthosis capable of measuring the angle of flexion of the fingers. By using support vector machines we could achieve an overall accuracy of more than 90%, with this methods being preferable to statistics and neural networks because of recognition accuracy and computational time.

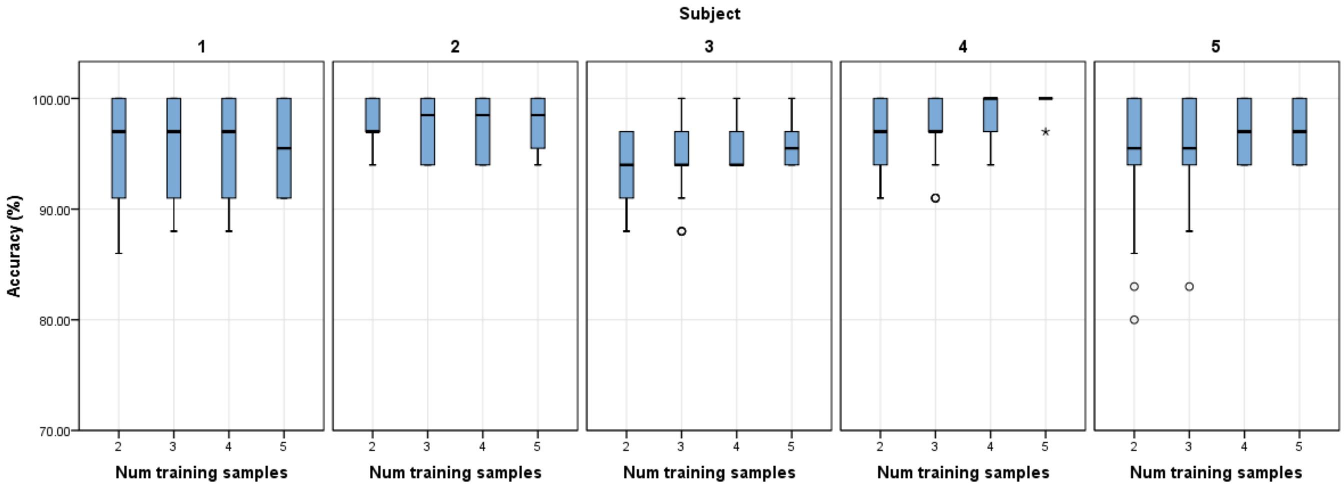


Fig. 8. Accuracy for each subject using the method based on SVMs.

This accuracy can be achieved already with only 4 training samples, thus allowing very short preparation time and making this approach suitable for home-based rehabilitation. In this sense, we also showed that the the fact that subjects pretending to grasp objects (instead of actually grasping them) had small effects on the recognition rate, suggesting that the proposed approach can work with imagined postures.

In future work, we will evaluate the performance of the selected method with data collected for post-stroke patients and compare the results with this study using healthy participants. Additionally, the set of grasps in this study was kept small in order to highlight the suitability of different algorithms, but the method could be tested including a larger number of grasps.

Acknowledgements

This work has been partly funded by the European Community Seventh Framework Programme under Grant agreement no. FP7-ICT-288698 (SCRIPT).

References

[1] A. A. Timmermans, H. A. Seelen, R. D. Willmann, and H. Kingma, "Technology-assisted training of arm-hand skills in stroke: concepts on reacquisition of motor control and therapist guidelines for rehabilitation technology design," J Neuroeng Rehabil, vol. 6, 2009, pp. 1–18.

[2] G. B. Prange, H. J. Hermens, J. Schafer, N. Nasr, G. Mountain, A. H. A. Stienen, and F. Amirabdollahian, "Script: Tele-robotics at home - functional architecture and clinical application," in 6th International Symposium on E-Health Services and Technologies (EHST), Geneva, Switzerland, 2012.

[3] J. P. Dewald and R. F. Beer, "Abnormal joint torque patterns in the paretic upper limb of subjects with hemiparesis," Muscle Nerve, vol. 24, no. 2, Feb 2001, pp. 273–283.

[4] T. Feix, H. bodo Schmiedmayer, J. Romero, and D. Kragic, "A comprehensive grasp taxonomy," in Robotics, Science and systems conference: Workshop on understanding the Human Hand for advancing robotic manipulation, 2009, pp. 1–2.

[5] A. Chaudhary, J. L. Raheja, K. Das, and S. Raheja, "Intelligent approaches to interact with machines using hand gesture recognition in natural way: A survey," International Journal of Computer Science and Engineering Survey (IJCSES), vol. 2, no. 1, 2011, pp. 122–133.

[6] L. Lamberti and F. Camastra, "Handy: A real-time three color glove-based gesture recognizer with learning vector quantization," Expert Systems with Applications, vol. 39, no. 12, 2012, pp. 10 489 – 10 494.

[7] J. LaViola, "A survey of hand posture and gesture recognition techniques and technology," Brown University, Tech. Rep. CS-99-11, April June 1999.

[8] D. Xu, "A neural network approach for hand gesture recognition in virtual reality driving training system of spg," in Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, vol. 3, 2006, pp. 519–522.

[9] R. Palm and B. Iliev, "Grasp recognition by time-clustering, fuzzy modeling, and hidden markov models (hmm) - a comparative study," in Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on, 2008, pp. 599–605.

[10] C. Systems, "CyberGlove," http://www.cyberglovesystems.com/, [retrived: January, 2014].

[11] S. Ates, J. Lobo-Prat, P. Lammertse, H. van der Kooij, and A. Stienen, "Script passive orthosis: Design and technical evaluation of the wrist and hand orthosis for rehabilitation training at home," in ICORR13, 13th International Conference on Rehabilitation Robotics, Seattle, Washington, 2013, pp. 1–6.

[12] S. Electronics, "Flex Sensor 2.2"," https://www.sparkfun.com/products/10264, [retrived: January, 2014].

[13] A. D. Keller, C. L. Taylor, and V. Zahm, Studies to determine the functional requirements for hand and arm prosthesis. Dept. of Engineering, University of California, 1947.

[14] N. Yozbatiran, L. Der-Yeghiaian, and S. C. Cramer, "A standardized approach to performing the action research arm test," Neurorehabil Neural Repair, vol. 22, no. 1, 2008, pp. 78–90.

[15] J. A. Anderson, An introduction to neural networks. Cambridge, MA: MIT Press, 1995.

[16] P. Melin and O. Castillo, "Supervised learning neural networks," in Hybrid Intelligent Systems for Pattern Recognition Using Soft Computing, ser. Studies in Fuzziness and Soft Computing. Springer Berlin Heidelberg, 2005, vol. 172, pp. 55–83.

[17] H. Hasan and S. Abdul-Kareem, "Static hand gesture recognition using neural networks," Artificial Intelligence Review, 2012, pp. 1–35.

[18] "Neurolab Library," https://code.google.com/p/neurolab/, [retrived: January, 2014].

[19] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, 1995, pp. 273–297.

[20] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, May 2011, pp. 27:1–27:27.

Needs and Usability Assessment of a New User Interface for Lower Extremity Medical Exoskeleton Robots

Yoon Jung Jeong
yjeong@me.berkeley.edu

Euiyoung Kim
euiyoungkim@berkeley.edu

Homayoon Kazerooni
kazerooni@berkeley.edu

Mechanical Engineering
University of California at Berkeley
Berkeley, CA USA

***Abstract*— This paper presents an evaluation and recommendations for the improvement of the user interface (UI) of medical exoskeleton robots for people with mobility disorders. Existing UIs of currently available medical exoskeletons lack the flexibility to serve a diverse user group who require more customization. A UI prototype consisting of a glove with buttons attached on fingertips, and a display module for user feedback and/or instruction was developed and evaluated. For the evaluation of this UI prototype, multiple usability tests, guerrilla tests, interviews, and surveys were conducted with several crutch and manual wheelchair users. Finally, a set of final Glove UI design recommendations is illustrated based on the test subjects and interviewees' feedback; finger glove, two buttons, singleton walking method, and adjustable display position. A more thorough evaluation on this improved UI with more potential users of medical exoskeletons with various physical abilities remains as future work.**

Keywords-Design for people with disabilities; User interface design; Exoskeleton; Glove Interface

I. Introduction

By the numbers, there are 700,000 new stroke patients every year, and 265,000 individuals suffering from spinal cord injuries in the United States [1]. These factors are the leading causes of mobility disabilities, which lead thousands of patients to use a wheelchair as their main mobility aid for the rest of their lives. However, due to the nature of a human body, the long-term usage of a wheelchair with prolonged sitting and immobilization of the limbs brings secondary diseases. Common examples of these secondary injuries are urinary tract infections, blood clots, reduction in cardiovascular functioning, decreased bone mineral density, and osteoporosis [2]. It is widely accepted in the field that postponing the use of wheelchairs will delay the onset of secondary injuries and diseases [3], [4], [5], [6]. It has been also shown that being upright and ambulatory has significant benefits to the human body, such as increased circulation, improved bladder and bowel functioning, and an overall feeling of well-being [2], [7].

Lower extremity medical exoskeleton technology can be beneficial for many wheelchair users. Through its ability to enable walking and standing upright, this technology can prevent secondary injuries as well as provide independence in the daily lives of those patients.

This paper presents the background of medical exoskeleton technologies, operating schemes, development and evaluation of UI prototype, result and recommendation for future work.

II. Background of Medical Exoskelteon Technologies and Overview of Their UIs

A. State-of-the-art medical exoskeletons

Powered medical exoskeleton devices that enable a user to stand and walk are available on the market with the price range of $100,000 - $150,000. There are Rex from Rex Bionics in New Zealand, ReWalk™ by Argo Medical Technologies in Israel, and Ekso™, developed by Ekso Bionics. These mobile, powered medical exoskeletons are made for rehabilitation or for walking outside the clinic. They have 4 to 10 powered joints aligned to the patient's biological joints. These powered joints are pre-programmed to mimic the human gait, which ambulate the wearer in a similar manner of natural walking when the user commands. Rex has a joystick-type UI located on the arm rest. Rex does not require crutches or a walker since it has self-balancing structure with powered ankles, knees, and hips [8]. ReWalk™ has powered knees and hips, and executes steps by sensing the tilt of the user's torso [9]. Ekso™ also has powered knees and hips, and has an operating method, which determines user intent using multiple sensors on the pilot's arms and crutches to estimate the user's pose and executes steps, as well as manual actuating via buttons on the crutches or walker for initial trainings [10]. AUSTIN, developed in Robotics and Human Engineering Lab in University of California at Berkeley, has a hip-knee coupled gait generation mechanism derived by powered hip joints, operated by a button-integrated UI on the crutch or walker handle [11].

B. User interface of medical exoskeletons

These state-of-art exoskeleton robots for patients with mobility disorders have two kinds of user input receiving systems; 1) sensing user intent from the wearer's posture, 2)

 ISBN: 978-1-61208-325-4

getting explicit commands from the wearer via joysticks or buttons. The first method has an advantage of requiring minimal effort from the user to issue commands. However, due to the limited mobility and muscle control of this user segment, this implicit UI has the potential risk of misinterpreting the operator's posture, which may cause falling. Studies have shown that the *implicit* UI of ReWalk™ results in a unique pattern of control for walking in individuals with spinal cord injuries, while patients with lower level injuries have better walking performances and also progress rapidly [12]. Having *explicit* UIs such as the systems used in Rex and AUSTIN requires more user effort in commanding steps via joysticks or buttons. However, this appears to minimize the risk of mistriggering due to the pilot's incomplete motor control due to the injuries. However, their device-coupled UI design (see Fig. 1) results in limited usability when used among different users with different hand/finger sizes.

In this paper, a glove type user-coupled human-machine interface prototype with tactile and visual feedback is discussed for a next generation *explicit* UI for medical exoskeletons. We present an evaluation of this UI prototype, including interviews, guerilla tests, surveys, and usability testing on people with mobility disorder. The evaluation preparation, methodologies, and results are discussed.

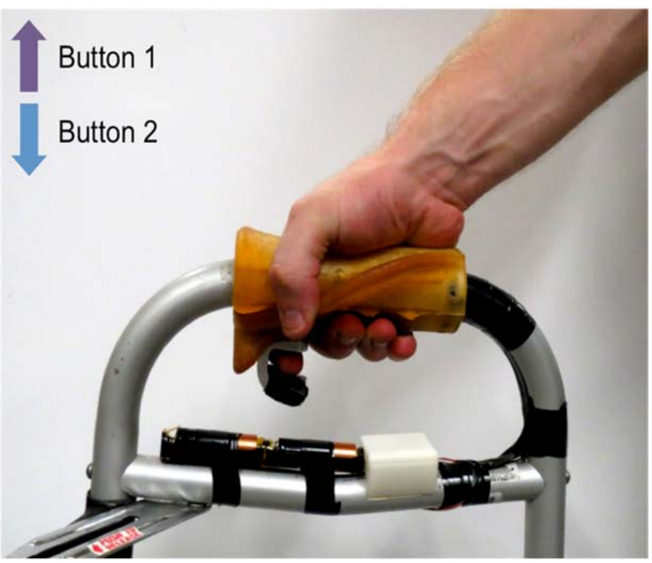


Figure 1. An example of device-coupled UI attached on a walker. The user can trigger the button upward or downward for two different commands. The comfort of this UI is highly dependent on the users' hand size.

III. Schematics of Explicit User Interface For Medical Exoskeletons

The explicit UI of medical exoskeletons switches the state of the machine from one state to the other when the user issues commands via buttons. There are four different states of the medical exoskeleton; left leg forward, right leg forward, feet together (standing up), and seated down.

With the walking scheme shown in Fig. 2, it is clear that at least two buttons - the simplest way of giving two different signals - are required for the user to switch from a state to the other state. For example, when the user is wearing an exoskeleton with his feet together, he/she can either press button #1 to start walking by moving his/her left leg forward, or click button #2 to sit down. Here, it is assumed that the user can sit down only when his/her feet are together, and the user only goes to "feet together" state from "seated down" state, not to left or right leg forward state. Obviously, there is more than just one way of mapping these button functions, and more than one implementation for the number of buttons the UI can have. One of the goals of this study is to find out the potential users' most preferred mapping which will contribute to the decision on long-term adoption of the device along with the hardware design of the UI.

For this study, two different walking operation schemes, i.e., methods of mapping buttons are suggested and evaluated: *Singleton* and *Alternating*. In the Singleton method (see Fig. 3 (a)), the user presses the same button for each step; for example, the user presses button A for the first step → user presses button A again to make the second step → (continue) → user presses button B to stand up (feet together) → user presses button B again to sit down.

The alternating method (see Fig. 3 (b)) designates different buttons for stepping forward with different legs. For example, in an alternating scheme, the user presses button A for the left step

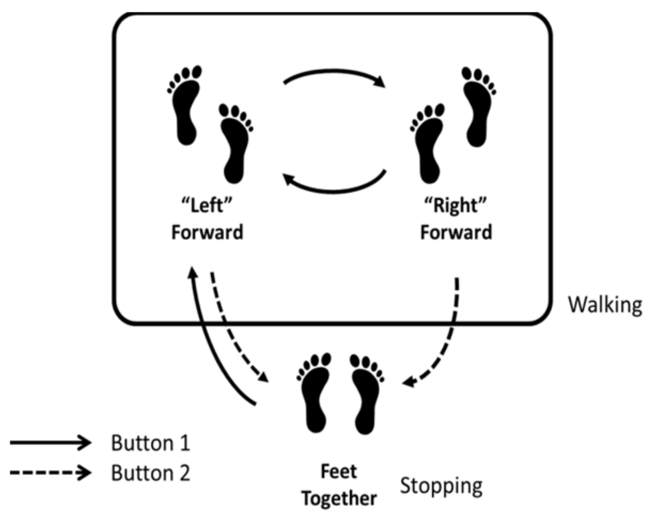


Figure 2. An example of a walking scheme. In this scheme above, the user stands up and starts walking by pressing button 1, and ends operation by pressing button 2.

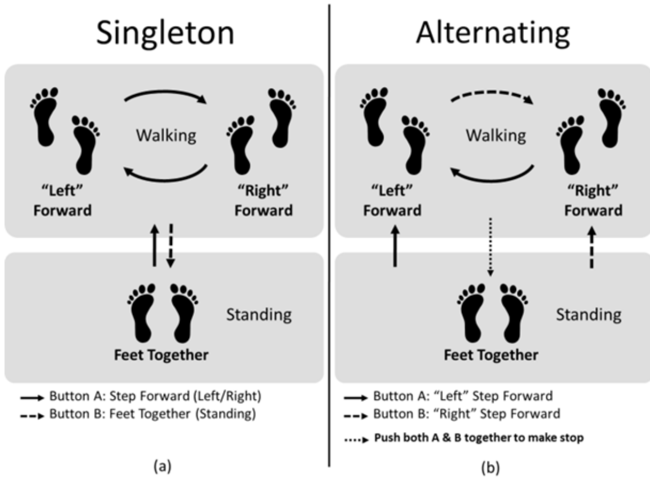


Figure 3. Two different walking schemes. (a) The Singleton method uses the same button for steping forward (b) The Alternating method maps two different buttons for stepping each leg forward.

→ presses button B to make the right step → button A to make the left step again → (continue) → user presses both buttons two times in a row to bring the feet together and sit down.

IV. PROTOTYPING

Several versions of prototypes have been developed to simulate the UI and elicit feedback from potential users during this testing phase. A "glove UI" featuring a glove with buttons attached on fingertips and an LED display mounted on the inside of the wrist (See Fig. 4) was prototyped as a suggested next generation UI design for medical exoskeletons.

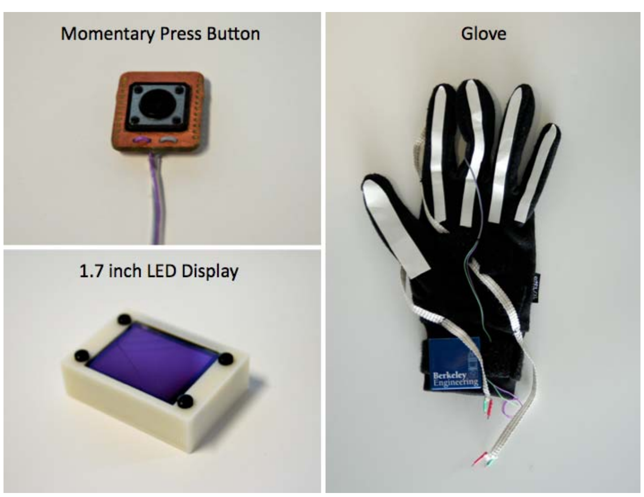


Figure 4. The three components of the glove user interface

Fig. 5 shows a wheelchair user wearing the refined prototype after pre-testing.

This new UI concept is designed to overcome the shortcomings of the other user interfaces of the state-of-art medical exoskeletons, including accommodating different users' hand sizes and gripping preferences, and providing visual feedback to the user via a monitoring display.

Figure 5. A subject wearing the prototype.

Information shown on the monitor includes On/Off status, battery life, and quick instructions for the exoskeleton operation.

V. RESEARCH METHODOLOGIES

A. Summary of Research Methodologies

Our data are collected from pre-testing, interviews, guerilla testing, surveys, and usability tests with eight subjects on our target user segment.

B. Pre-testing

The first prototype of the glove UI was tested at Berkeley Robotics and Human Engineering Lab by their medical exoskeleton test pilots. This pre-testing was used to revise our interview and usability test strategies according to the feedback from actual exoskeleton test pilots prior to testing the prototype with potential users who have never been exposed to this technology. While conducting a pre-testing, we found several improvements from our design. Given the distinctiveness of the potential user group, feedback from actual test pilots from Robotics and Human Engineering Lab was immensely valuable.

1) The UI prototype Improvement

While testing the prototype, we realized that is difficult to discern the hand position of a user who was wearing a black glove, and since a big part of our testing was determining the user's hand position, this was an important oversight. As a result, we taped white stripes to the back of each finger to make them more visible. We added Velcro to the back of the monitoring display so users could reposition the display. Letting the user determine their own ideal location was better than just asking if the display was readable.

2) The Interview/User Test Setup Improvment

Interviews with microphone and camera setups were conducted. We added a set of questionnaires to obtain the demographic data of test participants. We added a high-resolution camera (DSLR) to get detailed close-ups of motions while the subjects were testing out our prototype. We also used a microphone to closely record their voice and filter out the background noise.

Throughout the pre-testing, the subjects provided valuable feedback about our test environment and test questions/protocols, which ultimately helped the research preparations in the next steps.

C. Interviews/Guerilla Testing/Surveys

In order to collect information from our potential user segments, we conducted interviews at several locations where people with mobility disorders frequently visit. One of these places is the Ed Roberts Campus and the Center for Independent living located in Berkeley, California. Both guerilla and scheduled interviews were conducted on a wider variety of potential users at this center. A short list of survey questionnaires was also prepared to obtain basic demographic data from the participants before the interviews. Given the relatively small sample size, it was critical to coordinate in-depth interviews immediately followed by the usability test to thoroughly explore opinions on the glove UI prototype.

D. Usability Test

Usability tests were performed on three test subjects with mobility disorder – two crutch users (referred as Subject 1 and 3), and one manual wheelchair user (referred as Subject 2). The goal of this usability test is to find their preferences on the types of glove, the number of buttons, the location of the display, and walking scheme. A brief profile of the test subjects follow:

Subject 1: 30-year-old male, a relatively new crutch user due to his recent injury on his ankle.

Subject 2: 24-year-old male, paralyzed in his lower torso and legs due to car accident five years ago. He has been a manual wheelchair user for last five years.

Subject 3: 34-year-old female, a lifetime cane or single crutch user due to congenital leg abnormalities.

For usability testing, the subjects were given a glove UI prototype shown in Fig. 4, and asked to demonstrate how they would walk with their mobility aid while wearing the glove UI.

1) Grip, Buttons, and Glove Preference

The way patients hold their crutches or walker grips varies widely with hand size, type of mobility aid, and type of grip. Depending on these factors, the user's preference for button locations may differ for ease of use and mistrigger prevention. In testing preferences, subjects were asked to hold the grip of their mobility aid while wearing the glove UI prototype, and find the preferred buttons among the ones attached on fingertips of the glove UI prototype. The subjects were also asked for their opinions on the number and types of buttons. They were asked to choose at least two buttons for the simplest operation, and more buttons for other possible functions or more elaborated exoskeleton maneuver, if desired.

2) Type of Glove

A glove used for a prototype shown in Fig. 4 was provided in this usability test. The subjects were asked for their opinions on the type of glove in terms of the glove material; whether the fabric is too thick or thin, and whether the coverage of the glove over the hand is too much or too little.

3) Display Module Position

The display module has an LED screen, which is capable of displaying various types of information such as the system status, battery life, wireless connectivity, and error diagnostics when applicable. The subjects were asked to find the position of the display module where they could get information at a glance while they operate the exoskeleton. Our assumption here was that the preferred location of the display module would differ according to the type of mobility aid and its grip.

4) UI Scheme, or Walking Method Preference

Two different UI schemes described in section III were explained to the subjects, and the subjects were asked to simulate walking with the two different schemes, using their button preferences from the previous part of usability testing. The subjects were asked to follow each walking schemes in Fig. 4 for three times. After this simulation, the subjects were asked to describe their preferences and opinions on the two different schemes.

VI. Results and Design Recommendations

Overall, the results of the interviews and usability test appear to reveal the potential user's general preferences on medical exoskeleton UIs. Based on the feedback from the user research participants, design suggestions on four different components of the next generation glove type UI are summarized:

A. Type of Gloves

According to the preferences of usability test subjects and interviewees, adopting a single type of full glove for this application in meeting different kinds of user needs appears to be challenging due to the conflicting preferences among potential users. Manual wheelchair users responded that they generally prefer thick and durable gloves to protect their hands when they use a wheelchair. While a medical exoskeleton is to provide an alternative mobility aid, a wheelchair is still frequently used with medical exoskeleton during trainings and their daily life for easy maneuver and/or long travel. Therefore it appears that providing durable gloves for hand protection in wheelchair use as the glove UI platform would ease the don and doff of exoskeleton use. However, it appears that a full glove itself, even with thin material is not favored by other user groups since it provides unnecessary coverage, which might cause unpleasant heating on the user's hand. Therefore, a finger glove is recommended instead of a full glove. A finger glove satisfies the needs of non-wheelchair users, requires minimal hardware, and offers the option of wearing the glove UI on top of wheelchair users' thick gloves.

B. Number of Buttons

All the subjects responded that they liked the tactile feedback of the momentary switches on the prototype. Given the number of states of the machine and the users' desired operation, at least two buttons are required as stated in section III. For more elaborate maneuvering of the machine, using three buttons can be considered. However, it appears that having two buttons is preferred by the potential users due to its simplicity.

C. Walking Operation Scheme

The Singleton method is the preferred walking operation scheme due to its simplicity in operation. A majority of test subjects perceived the Singleton method as more intuitive and easier to operate. The Alternating method was deemed complicated for beginners even though it provides more flexibility. An argument remains to keep the Alternating method as an option for more experienced users. Key quotations from the subjects on these two walking schemes are following;

"I'm thinking about a computer game. Some people prefer using fewer but more intuitive commands, but I like to use more granular ones. But for right now, I like the Singleton option since I am new to this system."

"After a while if it was something that I had to do regularly, like, say, drive a car, it would probably just become second nature because I would train myself to click only one."

D. Display Module Position

Subjects showed different preferences regarding the display module's angle and position. Walker, crutch, and wheelchair users view the display from different angles due to their different grips. Therefore, it is recommended that the display module position stays adjustable with easy attachment methods such as the use of Velcro, for example.

VII. Conclusion

The goal of our research is to evaluate a new glove type UI for medical exoskeletons. A conceptual prototype that contains key components for evaluation was created and evaluated by potential users. Four research methods are used in this research; interviews, guerilla tests, surveys, and usability tests.

Finger glove, two buttons, singleton walking method, and adjustable display position are the final design recommendation based on our research.

- "Finger glove" instead of a whole glove
  This type of glove better facilitates users' different hand sizes, grip preferences, and the glove material.
- Adjustable display position
  Users with different types of walking aid have different preferences on the display module position.
- Two buttons rather than more buttons
  At least two buttons are required for an exoskeleton to provide a simple walking function. Having more buttons would provide more functions. However, potential users would prefer to start learning how to walk with exoskeletons with a minimal and simple interface.
- Singleton walking operation (see Fig. 3 (a))
  Mapping button functions related to the exoskeleton operation rather than the stepping leg appears to more simple and intuitive to use.

TABLE I. SUMMARY OF GLOVE UI DESIGN RECOMMENDATION

Components	Design choices provided to test subjects			
1. Type of Glove	 **Finger Glove**[a]	Vs.	 Whole Glove	
2. Number of Buttons	 **Two Buttons**[a]	***Vs***	 Three Buttons	 One Button
3. Walking Method	**Singleton**[a]	***Vs***	Alternating	
4. Display Position	 **Adjustable**[a]	***Vs***	 Fixed	

a. Highlighted features are the final design decisions

Final design recommendations are illustrated in Table I.

VIII. Discussion/Future Work

This study addresses the next generation UI design for medical exoskeletons that can potentially improve their usability and ability to learn. However, current state of the art medical exoskeleton technologies, as well as the outcome of this study, are limited to people who retain a certain level of upper body control, or can use crutches or a walker. Identifying the primary target user groups based on the level of their mobility and physical strength is critical regarding this user group that has a broad range of physical abilities that may affect various aspects of conceptual UI design. Exploring more customizable features to satisfy the needs of a broader range of users, for example, patients with higher-level injuries, remains as future work, along with testing with more test subjects.

 ISBN: 978-1-61208-325-4

We hope that the broader range of users with mobility disorders, other researchers, and practitioners in this field of research can benefit from our recommendations in the near future.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the test subjects who participated this usability testing and provided invaluable feedback. The authors also would like to thank interviewees from guerilla interviews from the Ed Roberts Campus and the Center for Independent Living, Berkeley, CA. Finally, the authors would like to acknowledge the Berkeley Robotics and Human Engineering Lab in the Mechanical Engineering department at UC Berkeley for providing opportunity to prototype and test the new UI design with their test subjects.

REFERENCES

[1] "National Spinal Cord Injury Statistical Center: Annual Statistical Report 2012," 2012. [Online, retrieved: January, 2014] https://www.nscisc.uab.edu/.

[2] D. Sabo, S. Blaich, W. Wenz, M. Hohmann, M. Loew, and H. J. Gerner, "Osteoporosis in patients with paralysis after spinal cord injury," Archives of Orthopaedic and Trauma Surgery, vol. 121, no. 1–2, Jan. 2001, pp. 75–78.

[3] DHHS –Office of Disability, Aging and Long Term Care Policy [Online, retrived; January, 2014] http://aspe.hhs.gov/daltcp/reports.htm

[4] Disability Statistics Center: University of California at San Francisco. [Online, retriveed: January, 2014] http://dsc.ucsf.edu/

[5] B. Kemp, "Quality of Life While Aging with a Disability", Assistive Technology, vol. 11, Feb. 1999, RESNA Publication.

[6] National Council on the Aging [Online, retrieved: January, 2014] www.ncoa.org/

[7] J. J. Eng, S. M. Levins, A. F. Townson, D. Mah-Jones, J. Bremner, and G. Huston, "Use of prolonged standing for individuals with spinal cord injuries.," Physical therapy, vol. 81, no. 8, Aug. 2001, pp. 1392–9.

[8] R. Little and R. Irving, "MOBILITY AID," U.S. Patent WO/2009/08224903-Jul-2009.

[9] A. Goffer, "Gait-locomotor apparatus," U.S. Patent EP126020119-Feb-2002.

[10] K. A. Strausser, T. A. Swift, A. B. Zoss, H. Kazerooni, and B. C. Bennett, "Mobile Exoskeleton for Spinal Cord Injury: Development and Testing," ASME Conference Proceedings, vol. 2011, no. 54761, Oct. 2011, pp. 419–425.

[11] H. Kazerooni et al., "Orthesis system and methods for control of exoskeletons", U.S. Patent US20130158445

[12] G. Zeilig, H. Weingarden, M. Zwecker, I. Dudkiewicz, A. Bloch, and A. Esquenazi, "Safety and tolerance of the Rewalk exoskeleton suit for ambulation by people with complete spinal cord injury: A pilot study," J Spinal Cord Med, vol. 35, no. 2, Mar. 2012, pp. 96–101.

Interactive Engagment Capabilities as an Indicator of E-Learning Systems' Usability

Aljawharah Al-Muaythir, Khalid Alnafjan, Lamia Al-Kwai
Software Engineering Department, Collage of Computer Science, King Saud University
Riyadh, Saudi Arabia
aljoharah1408@windowslive.com, kalnafjan@ksu.edu.sa, lalkwai@kacst.edu.sa

***Abstract*—Interactive engagement is easier to achieve in traditional learning environments (where face-to-face interaction is assured) than in e-learning environments. Therefore, a set of functionalities should be supported in e-learning systems to allow an acceptable level of interactive engagement, such as whiteboards, chartrooms, discussion boards, etc. In this study, some analysis and evaluation was conducted for a number of open source e-learning systems regarding their support for functionalities that aids in creating an interactively engaging learning environment. The evaluation included ATutor, Claroline, Dokeos, Integrated Learning, Information and Work Cooperation System (ILIAS), Modular Object-Oriented Dynamic Learning Environment (Moodle), Online Learning and Training (OLAT), and Sakai. The evaluation result showed that Moodle and Dokeos achieved the best coverage of all possible interactive engagement-supporting capabilities available, thus proving its superiority over other e-learning systems included in this study in providing an interactively engaging learning environment.**

Keywords— Interactive Engagment; Usability; E-learning; Open source.

I. INTRODUCTION

Educational Systems are evolving to meet the society's needs for learning systems that save time, money and the environment. E-learning systems, the use of information and communication technologies in the learning process, are emerging to meet those needs. They are the best solution for students and instructors that live in separate locations and for those who have temporal difficulties. Many benefits of e-learning systems are discussed in the literature, such as the flexibility of the material and time that is given to students, and more opportunities to ask questions and express thoughts. Also, it reduces the dependence on time constraints for teacher/lecturer and supports the accessibility to the course materials based on student's election [1]. In addition, J. Capper [2] listed several valuable benefits, including the fact that new learning strategies and approaches become economically feasible through e-learning systems. For instance, it becomes feasible to utilize faculty anywhere in the world and to put together faculty teams that include master teachers, researchers, scientists, and experienced professional developers. Also, group collaboration can be achieved by means of electronic messaging and shared conversations tools that give the participants the opportunity to work together as groups despite their physical locations.

The success of e-learning systems had led to the development of many e-learning platforms -either open-source or close-source- that support various tools and technologies to allow a good blend of learning capabilities in order to optimize the learning process. E-learning may be synchronous, where the learning among all participants occurs in real time, or asynchronous, where participants learn at there own pace. Weather the e-learning system was designed to be synchronous or asynchronous, participants are usually at various locations. Therefore, the assumption that e-learning systems are recreation of traditional systems is wrong due to the lack of face-to-face communication, body language, social cues in e-learning, but are essential in producing an engaging learning environment. Actually, traditional classrooms allow instructors to be aware of the level of student's understanding and easily employ some techniques to keep the students active and engaged. However, in virtual classrooms, that will be more difficult because of the lack of face-to-face interactions among students and instructors. In this situation, instructors must have access to some tools offered by the e-learning platform to promote the learning process [3]. Consequently, this allows instructors to make sure that students understand well, and keep them engaged by using several forms of interaction including: the interaction of students with the instructor, their fellow students, or the content. Instructors expect to use E-learning platforms that have adequate tools -whiteboards, audio and video are some examples of such tools that can be used in this context- to support the interactive engagement in virtual classroom.

Because of the remote nature of the e-learning activities, usability is of special importance in e-learning systems [4]. Furthermore, the level of interactive engagement supported in e-learning systems is found to be vital for achieving effectiveness [5], which is a core measurement of usability based on its definition in the ISO (International standard organization) 9241 standard-: "*The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use*".

Due to the importance of the supportability of interactive engagement in e-learning systems, this paper will emphasize how interactive engagement functionalities supported by e-learning systems affect their achieved usability. Furthermore, seven widely used open-source E-learning systems (Moodle [6], ILIAS [7], Sakai [8], ATutor[9], Claroline [10], Dokeos[11], and OLAT[12]) will be evaluated based on their interactive engagement capabilities.The comparison will be based on the offered synchronous and asynchronous tools by each one of the seven systems to support and meet the interactive engagement requirement. First, we will list and describe the tools that can be used generally in virtual classroom to support interactive engagement. Then, we will check the availability of these tools in each one of the e-learning systems included in this study. Finally, we will conclude our research with the usability evaluation –from the interactive engagement point of view- of these e-learning systems. In the next section we discuss the importance of interactive engagement in assessing the usability of e-learning systems, and in pedagogy. After that, we provide an overview of e-learning systems in general, and the e-learning systems that will be covered by this study. Following that, we will explain the applied evaluation approach. Then, we will list and explain the functionalities that support interactive engagement in e-learning systems, and on which we will base our comparison upon. Finally, we will present the evaluation results and conclude with our findings.

II. INTERACTIVE ENGAGMGNET IMPORTANCE

Hake [13] defines interactive engagement methods as ‘‘*those designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors*”, placing an emphasis on challenging the student to ‘think’ actively, which ensures that students pay and keep attention to the instructor and the material that is being presented. In this paper, we are interested in interactive engagement from two aspects: its importance in assessing the effectiveness (as a direct indicator of systems’ usability) of e-learning systems, and its importance in pedagogy to achieve educational goals.

A. Interactive Engagement Importance in Usability Assessment of E-learning systems

Interactive engagement is one of the important needs by users (learners and instructors) that must be considered and met to produce an e-learning system with high usability. In Michael Allen's guide to e-learning [28], interactive engagement capabilities is one of the fundamental services that should be provided in e-learning systems. Furthermore, one of the three shifts that should be adopted by e-learning describes in this guide is the shift from passive learning to interactively engaging learning [29].

From usability point of view, each system has a set of user profiles and each should be able to achieve a number of tasks properly. Learners and instructors are the common user profiles of e-learning systems and should be able to perform tasks through offered functionalities in order to fulfill the interactive engagement user requirement [17].

In addition, learners and instructors have the need to use collaborative and interactive tools to share their tasks and ideas with each other [18]. Furthermore, the use of interactive engagement materials and tools significantly enhances the user experience in e-learning [30]. Given the importance of these requirements (interactive engagement and collaboration functionalities), to achieve an effective learning process through the use of e-learning systems, the evaluation of e-learning usability based on the design heuristics is not enough. Thus, users may abandon an e-learning system if the interactive engagement –as an essential pedagogy- is not supported.

Some previous studies [19] [20] have established a framework to evaluate the usability of e-learning systems, which combine the instructional design and usability heuristics. Another, [17], has proposed a systematic approach based on user requirements and goals.

One of the important studies that considered the interactive engagement as one of the usability parameters in their proposed evaluation framework is [21]. This framework is based on learner-centered-perspective and combined usability and instructional design parameters, such as learnability, consistency, multimedia use, and interactive engagement. Also, for each parameter, there is a group of measurement criteria, such as the use of games and media as measurement criteria for interactive engagement parameter.

These previous studies indicate that the usability of e-learning systems can be affected by the pedagogies and learning goals that should be placed under great importance, such as interactive engagement, and the need to combine the pedagogical guidelines with design heuristics to create effective usability evaluation frameworks for e-learning systems.

B. Interactive Engagement Importance in Pedagogy

Communication and interactions between instructors and students are essential for the quality of learning. Interaction in pedagogy between instructors and students applies to any form of communication, whether initiated by the instructor or by the student. The four types of interaction that may occur in education are learner-to-content, learner-to-learner, learner-to-technology, and learner-to-teacher [14]. The occurrence of these types of interactions is essential in creating an interactive engaging learner environment and diminishing the learning barriers. Many forms of communication may be applied in order to achieve an interactive and engaging learning environment, including:

- Discussions between the instructor and the student, a group of students, or within groups of students.
- Student Expressions and thoughts about presented topics.
- Feedback provided from instructors on students discussion, academic performance, etc.
- Group activities.
- Applied learning strategies such as interactive video, audio, questionnaires, games, etc.

Educators and pedagogical researchers have placed great importance in engaging the learner; and “To teach is to engage students in learning” [15] is a famous quote in pedagogy. Moreover, gains from interactively engaging

 ISBN: 978-1-61208-325-4

the learner vary from enhancing critical thinking, to reducing student attrition [16].

III. OVERVIEW OF E-LEARNING SYSTEMS

E-learning refers to the use of Information and communication technologies in education and learning. It is enabled through the use of e-learning systems, which offer a wide range of capabilities and features from simple asynchronous content management, to synchronous virtual classrooms. The most agreed upon terms used to classify such systems based on their capabilities and features are: Learning Management System (LMS), which are software applications that deliver, manage, and track e-learning education courses\ training programs, Content or Course Management System (CMS), which are software applications that manage –in terms of creation and administration-content on educational\training websites, Learning Content Management System (LCMS), which are software applications that fulfill the characteristics of a LMS (administrative and management) and a CMS (content creation and administration), and Virtual Learning Environment (VLE), which are software systems that provide virtual classrooms for students and teachers to interact with each other by the integration of web 2.0 tools.

The features and capabilities that are implemented in an e-learning system depend on specific institutional needs and targeted markets. Though, common features in most e-learning systems include [22]:

- Structure – organization of all learning-related functions into one system.
- Security – protection from unauthorized access to all system content.
- Registration – finding, selecting or assigning courses.
- Delivery – delivery of learning content to learners.
- Interaction – including some types of learner interaction: with the content, with other learners, with instructors, and with course administrators.
- Assessment –collection, tracking, and storing of assessment data.
- Tracking – tracking of learner data indicating progress\usage.
- Reporting – extraction and presentation of required information.
- Record keeping – storage and maintenance of data about learners.
- Personalization – configuration of e-learning system to match personal preferences\organizational needs.
- Integration – the ability to exchange data with external systems.
- Administration – centralized management of all system functions.

Due to the notable advantages of e-learning and the widespread application of information and communication technologies, many organizations and schools are making more investments in acquiring e-learning systems to support and improve learning [23]. Cost-effective alternatives to buy or develop in-house e-learning systems are open source e-learning systems. Currently, there are more than 250 commercial e-learning systems and more than 45 of them are Open Source Software (OSS) [24].

Among the most popular open source e-learning systems are Moodle, Dokeos, Claroline, ILIAS, OLAT and others.

To ensure a fair argument, this paper will compare seven e-learning systems under the same category -open source- namely ATutor, Claroline, Dokeos, ILIAS, Moodle, OLAT, and Sakai. The selection of open source over proprietary e-learning systems is because of the advantages they provide, such as [25]:

- Ease of customization to meet specific organizational needs because their code is open.
- Ease of localization by modifying language preferences.
- No licensing costs.
- Faster bug fixes because of active supporting communities.
- Compatibility and extensibility with third party add-ons.

Following is a brief description of each one of the e-learning systems under comparison.

A. *Moodle*

Moodle is a free, open source educational software platform categorized as an LMS and VLE. It was developed with a focus on helping educators create online courses that are interactive and collaborative. It was released to the public in 2002 and its high modularity and ease of use are of its most appealing advantages and is what makes it most popular among small-to-medium and higher education business markets.

B. *ILIAS*

ILIAS is a free, open source, web based LMS and VLE. Although the project started in 1997, it was only released for public in 2000 as open source software. ILIAS is commonly used in public and private institutions and companies because of its multi-purposes: course player tool, authoring tool, and communication and collaboration platform. It is also popular in the security and defense sector because of its defense-specific security and interoperability requirements considerations.

C. *Sakai*

Sakai is a free, open source educational software platform considered as an LMS, CMS and VLE. Sakai was built by a consortium of five large U.S. universities and is based on existing tools contributed by each of these universities. It was first released to the public in 2005 and has gained popularity for its scalability, security, and high support for end features, which made it especially popular among large universities.

D. *ATutor*

ATutor is a free, open source LCMS and VLE. ATutor was designed with emphasis on accessibility (access alternatives to screen elements, and text alternatives to visual screen elements) and adaptability (with various teaching and learning scenarios). ATutor was first released in 2002, and it is effective for small and large organizations presenting their materials on the Web, or delivering online courses.

E. Claroline

Claroline is a free open source LMS and LCMS that serves the educational more than the corporate sector. It provides a space for training and collaboration and was released for the public in 2001. Claroline is based on pedagogical principles found in literature. It is popular for its high adaptability to different training/learning contexts, customizability by offering a flexible work environment, and the valuable offered tools that enable the instructor to use it in order to create rich course material.

F. Dokeos

Dokeos is a free open source corporate learning suite, which is technically considered as an LMS, and VLE. It has started as a company and LMS in 2004. It has four separate components to build e-learning content, to handle interaction with learners, to sell a course catalog, and for assessment and certification. It is popular for its high simplicity and reporting capabilities. Also, it supports flexible free or non-free extended tools such as authoring, video conferencing, quiz building, assessment, and reporting to administration tools.

G. OLAT

OLAT is a free open-source learning platform that launched in 1999. It is considered as an LMS that offers several languages and features that support e-assessments, collaborative groups, and interactive/effective learning.

IV. Applied Evaluation Approach

Here, we will compare and evaluate seven most popular open source e-learning systems, described above, namely, ATutor, Claroline, Dokeos, ILIAS, Moodle, OLAT, and Sakai. This evaluation is conducted to check the available functionalities that may support and enhance interactive engagement in the learning process offered by these systems.

To evaluate such systems, we investigated eight synchronous and asynchronous capabilities that can aid to support interactive engagement. These capabilities are (virtual classroom, messaging, forums, chat room, wikis, groups, games, multimedia support).

Our evaluation consisted in installing such systems, and using them as regular users with the help of systems manuals/documentations (if needed). We used a score scale (0-3) for the evaluation; 0 if the system does not offer the capability, 1 if the capability is partially supported and system does not offer the adequate main functionalities, 2 if the system supports the capability by offering the main functionalities, 3 if the system provides more functionalities than the main ones, so the capability is highly supported by the system.

V. Interactive Engagement Capabilities

In this section, we will list and explain the capabilities that support interactive engagement in e-learning systems, and on which we will base our comparison between ATutor, Claroline, Dokeos, ILIAS, Moodle, OLAT, and Sakai.

Mainly, most of these capabilities-if supported by a system- are offered as plugins or by integration, so users can install and configure any one of these functionalities based on his needs. Communication, collaboration, and cooperation are very important factors to introduce interactive engagement. Based on that, we selected eight synchronous and asynchronous functionalities that support these three factors and may lead to obtain an interactively engaging learning environment. The selected capabilities are as follows:

1. Virtual classroom: it includes three main capabilities which are:
   - Whiteboard: synchronous tool that keeps the students focused during class by allowing them to clarify ideas, and share what they draw/write on it.
   - Web and video conferencing: synchronous tool allows users in separate multiple locations to create real time classroom by using telecommunication technologies to simultaneously communicate two-way video and audio transmissions, content sharing, and slideshow presentation.
   - Screen sharing: by depending on number of protocols, it is a feature that allows the user to connect to a computer in separate locations to see that computer's desktop, and interact with it as if it were local.
2. Messaging: asynchronous tool that allows students to send emails/messages to their fellow students or instructors and vice versa.
3. Forums: it includes:
   - News/announcement forum: asynchronous tool that allows users to announce about any event/activity related to courses.
   - Discussion forum: asynchronous tool for students and their instructor to discuss any ideas or details related to any course class/ lecture or issues.
4. Chat room: synchronous text based discussion.
5. Wikis: asynchronous tool used to create/update collaboratively web pages/documents by students and instructors.
6. Groups: asynchronous tool that allows instructor to divide students as groups to do a specific task or activity.
7. Game: this functionality is used to apply game based learning by offering some games that achieve a specific goal in the learning process.
8. Multimedia support: a main capability that support adding/embedding, sharing, and viewing video/audio/image files related to course/class content.

VI. Evaluation Results

In this section, we present the evaluation results of our study. Table 1 shows the results for each system based on the offered functionalities. Mainly, all systems offer a basic set of synchronous and asynchronous functionalities that help the instructor to increase the level of interactive engagement in a course.

The results in Table 1 show that Moodle and Dokeos achieve higher mean value than other systems, and that means they offer more functionalities related to interactive engagement, ILIAS and ATutor offer less functionalities than Moodle and Dokeos. Sakai and Claroline achieve the same value of interactive engagement support. Although these systems achieve different values of support, they provide the main and important interactive engagement capabilities. One of these important capabilities-virtual classrooms- is slightly supported by OLAT that achieve the lowest support value.

Although the scores that were achieved appear to be close, the impact of an additional functionality can increase the usability of the e-learning system greatly [26].

All e-learning system that have been evaluated, other than Moodle and Dokeos, do not support the game functionality, which is a valuable capability that assists to make the learning fun and helps to keep the students focused and engaged. This kind of capability supports game based learning that helps to increase the engagement of students [27].

Also, Sakai Claroline and OLAT do not highly support Multimedia services, which is a basic capability that is used by most e-learning systems to assist the learning process. Multimedia support capability can improve the interaction/communication between students and their instructors by enabling them to share some media related to their course, discuss it, and post comments.

Furthermore, OLAT does not provide enough virtual classroom functionalities, such as whiteboard and video conferencing, to support the interactivity of the learning process.

VII. CONCLUSION

In this paper, we showed that interactive engagement functionalities are essential in evaluating the usability of e-learning systems. The study first investigates the importance of interactive engagement from a pedagogical point of view, and its importance to be supported by e-learning systems to enhance their effectiveness and therefore their usability. After that, we suggested a list of eight capabilities that may provide an interactively engaging e-learning environment. Furthermore, we evaluated a number of open source e-learning systems to check their support for these functionalities. As a result, we found that Moodle and Dokeos can introduce better interactive engagement by supporting more functionalities than ILIAS, ATutor, Claroline, OLAT and Sakai.

Although the results of the evaluation were close, each single functionality adds great value to an e-learning system and its capability of providing an interactive engaging learning environment.

TABLE 1. EVALUATION RESULTS

System / Functionality	Moodle	ILIAS	Sakai	ATutor	Claroline	Dokeos	OLAT
Virtual classroom	3	3	3	3	3	3	1
Messaging	3	3	3	3	2	3	2
Forums	3	3	3	3	2	3	2
Chat Room	3	3	2	3	3	2	3
Wikis	3	3	3	3	3	3	3
Groups	2	3	2	3	3	3	3
Game	2	0	0	0	0	2	0
Multimedia support	3	3	2	3	2	3	2
Mean value	2.75	2.62	2.25	2.62	2.25	2.75	2

REFERENCES

[1] D. Bouhnik and T. Marcus, "Interaction in distance-learning courses", Journal of the American Society for Information Science and Technology, 57(3), 2006, pp. 299-305.

[2] J. Capper, "E-learning growth and promise for the developing world", TechKnowLogia, 2(2), 2001, pp. 7-10.

[3] V. Junk, N. Deringer, and W. Junk, "Techniques to engage the online learner", In Proceedings of Hawaii International Conference on Education, Honolulu, Hawaii, 6, Jan. 2007, pp. 1-27.

[4] S. K. B. Wong, T. T. Nguyen, E. Chang, and N. Jayaratna, "Usability metrics for e-learning", In On The Move to Meaningful Internet Systems 2003: OTM 2003 Workshops, Jan. 2003, pp. 235-252, Springer Berlin Heidelberg.

[5] I. Roffe, "E-learning: engagement, enhancement and execution", Quality Assurance in Education, 10(1), 2002, pp. 40-50.

[6] https://moodle.org/ (retreived: November, 2013).

[7] http://www.ilias.de/docu/goto.php?target=root_1 (retreived: November, 2013).

[8] https://sakaiproject.org (retreived: December, 2013).

[9] http://atutor.ca (retreived: November, 2013).

[10] http://www.claroline.net/short-presentation/?lang=en (retreived: December, 2013).

[11] http://www.dokeos.com (retreived: November, 2013).

[12] https://www.olat.uzh.ch/olat/dmz/ (retreived: November, 2013).

[13] R. R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses", American journal of Physics,1998, pp. 66, 64.

[14] M. G. Moore, "Editorial: Three types of interaction", 1989, pp. 1-7.

[15] C. R. Christensen, "Education for Judgment: The Artistry of Discussion Leadership", Harvard Business School Press, Boston, MA 02163, 1991.

[16] K. A. Smith, S. D. Sheppard, D. W. Johnson, and R. T. Johnson, "Pedagogies of engagement: Classroom-based practices", Journal of Engineering Education, 94(1), 2005, pp. 87-101.

[17] L. Triacca, D. Bolchini, L. Botturi, and A. Inversini, "MiLE: systematic usability evaluation for e-learning Web applications", In World Conference on Educational Multimedia, Hypermedia and Telecommunications Vol. 2004, No. 1, 2004, pp. 4398-4405.

[18] C. J. Bonk, "Online training in an online world", Bloomington, IN: CourseShare. Com, 2002.

[19] L. Benson et al., "Usability and instructional design heuristics for e-learning evaluation", In World Conference on Educational

Multimedia, Hypermedia and Telecommunications, Vol. 2002, No. 1, 2002, pp. 1615-1621.

[20] L. L. Lohr,and C. Eikleberry, "Learner-centered usability. Tools for creating a J learner-friendly instructional environment. Performance Improvement", 40(4), 2001, pp. 24-27.

[21] P. Zaharias, "Usability in the Context of e-Learning: A Framework Augmenting 'Traditional'Usability Constructs with Instructional Design and Motivation to Learn", International Journal of Technology and Human Interaction (IJTHI), 5(4), 2009, pp.37-59.

[22] P. Berking, and S. Gallagher, "Choosing a Learning Management System", 2012.

[23] C. C. Aydin, and G. Tirkes, "Open source learning management systems in e-learning and Moodle", In Education Engineering (EDUCON), April. 2010, pp. 593-600, IEEE.

[24] A. Al-Ajlan, and H. Zedan, "Why moodle", In Future Trends of Distributed Computing Systems, 2008. FTDCS'08. 12th IEEE International Workshop on, Oct. 2008, pp. 58-64, IEEE.

[25] Monarch Media Inc. Business white paper "open–source learning management systems:Sakai and Moodle", 2010, www.Monarchmedia.com (retreived: December, 2013).

[26] C. Ardito et al., "Usability of e-learning tools". In Proceedings of the working conference on Advanced visual interfaces, May. 2004, pp. 80-84, ACM.

[27] N. Hallinen, E. Walker, R. Wylie, A. Ogan, and C. Jones, "I was playing when I learned: A narrative game for French aspectual distinctions", In Proceedings of the Workshop on Intelligent Educational Games at the 14th International Conference on Artificial Intelligence in Education, Brighton, UK, 2009, pp. 117-120.

[28] M. W. Allen, "Michael Allen's guide to e-learning: Building interactive, fun, and effective learning programs for any company", Wiley. Com, 2003.

[29] G. Conole, "Describing learning activities. Rethinking pedagogy for a digital age", 81, 2007.

[30] S. Alexander, "E-learning developments and experiences", Education+ Training, 43(4/5), 2007, pp. 240-248.

Interactive Systems Adaptation Approaches: A Survey

Imen Jaouadi
MIRACL, FSEG
Sfax, Tunisia
jaouadiiimen@gmail.com

Raoudha Ben Djemaa
MIRACL, ISIMS
Sfax, Tunisia
Raoudha.Bendjemaa@isimsf.rnu.tn

Hanene Ben Abdallah
MIRACL, FSEG
Sfax, Tunisia
Hanene.BenAbdallah@fsegs.rnu.tn

***Abstract*— Nowadays, the design and the development of user interfaces impose new requirements as a result of the diversity of platform with specifics characteristic. In this context, several approaches are proposed to develop adaptable user interfaces to context of use. This paper presents a review and classification according to different criteria of the most important research efforts in this field. In the light of this analysis, we propose to develop in the future researches a approach-based on model which adapts the functionality and user interface of an interactive system at run time.**

Keywords-user interface; adaptation; context of use; HCI.

I. INTRODUCTION

The technological progress, such as the miniaturization of microprocessors and the sensors and the success of communicating technologies, opens a wide field of possibilities in the development of Human Computer Interface (HCI) [19]. The user of application wishes to have information whenever and wherever he/she is located. Multiple types of interactive applications are necessary to adapt to the user profile (novice, expert, children, etc.) taking into account technological advances, such as the development of new platforms. The design and the development of interactive systems impose new requirements. It is necessary to design interactive systems in an abstract way, because there is adversity of different platforms with specific features. In this context, several approaches have been proposed in literature to develop adaptable HCI to the context of use [6][12].

This paper reviews major contributions made in the development of adaptable interactive systems to the context of use. The review highlights aspects that are not addressed or no existent in previous research efforts. It concerns those aspects we believe will assume a greater level of importance for future User Interfaces (UIs).

The remainder of this paper is structured as follows. We will discuss relevant related work in Section 2. In Section 3, we will characterize approaches studied on the basis of criteria. Finally, we conclude and outline our future work.

II. RELATED WORK

In this section, we present representative examples of approaches aimed at the adaptation of interactive systems to the context of use. We can distinguish three families of approaches to adaptation: approaches-based on components, approaches-based on models and approaches coupling between models and components.

A. Approaches-based on Models

Thevenin [15] provides a conceptual framework formed by several steps. In this framework, he proposed to add three more models than proposed by Szekely [14], which are: the environment, the user and the platform. In addition, in his research, Thevenin [15] offers a demonstration of his work called: ArtStudio. However, this tool only supports the diversity of platforms. In addition, it does not allow dynamic adaptation. Indeed, adaptation is assisted by the user. Thus, it does not cover all the principles of conceptual framework.

Sottet [13] is considered as a pioneer to have proposed coupling the model driven Engineering and the human computer Engineering. In his approach, Sottet [13] offers Task metamodel, Concept metamodel, User Interface Abstract (UIA) metamodel, and UI Concrete (UIC) metamodel covering Camelon reference framework [3]. In fact, Sottet [13] proposed adaptation controlled by a decision system in order to automate the generation of plastic HCI. This system allows choosing the adequate transformation to a given context. Increasingly, Sottet [13] developed a demonstrator called MARA, which is based on metamodels and transformations. The major drawback of this approach is that we have to create N transformations for N contexts and the decision system selects the most appropriate transformation to a given situation.

Hachani et al. [10] propose a new generic approach for the generation of UI adaptable. The approach is based on the variability [9] and Model Driven Engineering (MDE) [7]. The concept of variability can factorize the common parts to all contextual situations by identifying the variable parts. In this approach, Hachani et al. [10] propose to introduce the context of use at task level. This approach is distinguished by the definition of the generic rules appropriate to all contexts of use. However, Hachani et al. [10] do not specify a detailed description of each context element. Thus, the approach focuses only on the modeling task and context, and does not take the other models proposed by Camelon reference framework [3].

Bouchelligua et al. in [2] propose an approach-based on IDM for plastic HCI. This approach allows the adaptation to the context of use (platform, environment and user) based on parameterized transformation [17]. To apply the transformation, Bouchelligua et al. include in their approaches two abstraction levels: UI Abstract and UI Concrete. Thus, Bouchelligua et al. provide the context metamodels used to adapt the interface: user metamodel, platform metamodel, and environment metamodel. This approach is distinguished by modeling context elements, but it only supports adaptation at UI Abstract and UI Concrete. Other levels of abstractions defined in Camelon reference framework are not considered in this approach. Thus, Bouchelligua et al. have not developed a tool for their approach.

B. Approaches-based on Components

In [1], Balme et al. propose a conceptual framework as a functional decomposition that collects and organizes all the functions necessary for plasticity. The necessary plasticity functions are divided between the *infrastructure* dedicated to capturing the context and *adaptation manager*. Interactive systems are represented by components assemblies and connectors. The adaptation manager decomposes into three functions: *situation identifier*, *adaptation producer*, and *evolution engine*. The situation identifier can observe, match, and synthesize situations. *Evolution engine* select the most suitable components for the current state. *Adaptation producer* responsible for the implementation of adaptation plans. *Component manger* allows the dynamic discovery of components. In addition, Balme et al. [1] have developed a demonstrator for their approach called "*Ethylene*".

The proposed approach has several advantages: it is context aware and allows the user to control the system. However, in this approach we are forced to use a predefined and fixed set of components which are pre developed for specific needs.

In [11], Hariri et al. describe a design process to generate the UI from an Abstract model and/or Task model preserving ergonomic properties of HCI. This process allows the dynamic adaptation in runtime depending on the context of use. They used patterns and integrating a software architecture based on business components. A business component is composed of functional components containing the functional core and presentation components corresponding to the HCI. In addition, this approach is based on the concept of learning, that allows to continue develop the knowledge base of the system at runtime. Thus, the major advantage of this approach permits to adapt the functional core. Nevertheless, this approach has some limitations. In fact, the proposed method is not supported by a global environment; no complete development tool has been developed for the creation and management of design patterns and business. Furthermore, Hariri et al. [11] have not studied the concept of HCI migration to another totally different modality; for example, the migration of a graphical platform to voice platform.

In his work [8], Gabillon is interested in HCI dynamic composition. It is focused on the composition of the Task model from the user goal. This approach uses planning algorithms to dynamically compose the Task model. The HCI are implemented by comets [5]. Each comet is described by an operator or a method of planning. During his work, Gabillon [8] develops a demonstrator of HCI composition by planning called "Compose". In this tool, the contexts of use and ergonomics criteria are modeled manually. The major drawback of this approach it based only on the Task model and the purpose of the user. Thus, it is necessary to automate the inclusion of ergonomic criteria to dynamically compose ergonomic HCI. In addition, the preferences of users are not taken into account.

C. Approaches-based on Components and Models

Comets approach [5] provides a model for plastic interactors that are able to adapt to the context of use. Adaptation is based precisely on the description of an interactor in terms of resources. Resources refer to the terms of needs interactor functionality such as screen space. In [5], Dâassi et al. propose to encapsulate in the same component all Camelon reference framework models [3] and adaptation mechanisms. However, we believe that encapsulate in the same software component (the comet) all model specifications and adaptation mechanisms surcharge component. In addition, adaptation aspect and self-adaptation at the base component may affect the ergonomic usability of the system. Increasingly, Dâassi et al. [5] have not developed a tool to support this approach.

Criado et al. in [4] present a MDE [7] approach to development of adaptable UIs. Indeed, in this approach the UI can regenerate themselves during execution depending on user interactions and application requirements. UI components can be evolved over time through transformations of models changeable and adaptable according to the system events. However, the proposed adaptation process is not automatic; transformations and events are launched manually. Thus, the approach allows only the adaptation of interface model. In addition, Criado et al. [4], propose only six types of actions that can be performed on the component: create, delete, activate, deactivate, interact service and launch event.

III. ANALYSIS OF THE STATE OF THE ART

Table 1 provides visual summation of characteristics criteria are useful to answer respectively to the questions *Adaptation moment*, *Context element*, and *Adaptation level* Vanderdoncket et al. in [18]. The other criteria are inspired (who), (when), (with respect to what), and (what) posed by used to perform comparison between the approaches presented in the state of art (Section I). *Adaptation type*, from the analysis of works presented in Section I and our knowledge in HCI field.

TABLE 1. COMPARAISON BETWEEN APPROCHES REPRESENTED IN RELATED WORK

Approach	**Approach type**	**Adaptation type**	**Adaptation moment**	**Context element**	**Adaptation level**	**Models/functions**	**Learning mechanisms**	**Associated tool**
Thevenin [15]	based on model	Assisted generation	design	platform	UI	Domain, task, UI Abstract UI Concrete, UI Final, platform, environment	No	ARTStudio
Commet [5]	based on model + component	dynamic	design	interaction context	UI	Domain, Task, UI Abstract UI Concrete, UI Final User, Platform Environment, Transition, evolution	No	Camnote ++
Balme [1]	based on component	dynamic	runtime	interaction context	UI	situation identifier, evolution engine component manager, adaptation producer	Meta-iu	Etylene
Sottet [13]	based on model	Statique dynamic	design	platform	UI	Domain, Task, UI Abstract UI Concrete, UI Final, user, platform, environment,quality	No	MARRa
Hachani [10]	based on model	static	design	interaction context	UI	task	No	No
Hariri [11]	Based on component	dynamic	runtime	interaction context	Functional core	Task, UI Abstract, UI Concrete, UI Final	knowledge base	No
Criado [4]	based on model + component	manual	runtime	No	UI	Structural metamodel, visual metamodel, interaction metamodel	No	No
Gabillon[8]	based on component	dynamic static	runtime	interaction context	UI	Domain, task, UI Abstract UI Concrete, UI Final	No	Compose
Bouchelligua [2]	based on model	manual	design	Interaction context	UI	UI Abstract, UI Concrete	No	No

In this article, we identify three main groups of adaptation approach. These are: approach-based on model, approach-based on components and approach-based on components and model.

We can see that the component approach solves the problem of dynamic adaptation. The major drawback of these approaches is that they are based primarily on the use of fixed and predefined set of variants; whereas, the model-based approach allows representing more variants. Therefore, we propose in our future work a based model approach to adapt the HCI to different contexts of use.

We can also arrange approaches depending on the time of adaptation (*Adaptation moment*). Many researchers in the reviewed literature are proposed to adapt HCI at the design phase [2][5][10][15]. Others focused on the HCI adaptation during execution [1][4][8]. It is apparent from the review that adapting HCIs during execution is almost appropriate for the approaches-based on components. In fact, in these approaches adaptation is performed by a components dynamic assembly, while the adaptation of model-based approaches during execution is a very difficult task. It forces the system to generate code for each new adaptation.

In the literature, several approaches are proposed to develop HCI adaptable to the context of use. The major of researchers are restricted to narrow classes of context such as platform [15] or language [10]. This one is due to the difficulty of capturing and processing contextual information. There are approaches that have considered all the contexts of use; even if they did not specify a detailed description of each component of the context of use [8][10]. These lacks do not make these approaches useful in a general context and especially not for applications where personalization information is part of the essential criteria of system operation.

According to [16], an interactive system is composed of four functions: *physical presentation*, *logical presentation, functional core adaptor,* and *dialogue controller*. The adaptation of the HCI can take place at all these levels. An adaptation of the physical presentation preserves the nature of physical interactors, but their rendering may be different for other platforms. The adaptation of logic presentation changes the nature of their interactors but not its representational and functional capacity. The Adaptation of dialogue controller changes task scheduling but not their nature. However, adaptation at the adapter functional core result of the changing nature of concepts and exported by the functional core functions. It changes the nature of the tasks and concepts they handle. It is apparent from the review that a lot of research [1][2][4][5][8][10][15] aim to manipulate user interface without concern about the functional part. From the reviewed approaches, only Hariri et al. [11] are concerned with adapting the functional core of HCI. It is based on components. In fact, to our knowledge, there are no model based approaches proposed in the literature which are adapted functional core and the user interface of the interactive system.

In addition, each approach puts forward a number of models to adapt the system to context of use. Thus we find that some approaches focus on the Task model like central model for adaptation [10]. Many also introduced in their approaches all models defined by Szekely in [14]: Domain, Task, UI Abstract, UI Concrete, and UI Final [5][8][11][15]. These approaches have shown how to gradually build different models to define all the characteristics necessary for the construction of the final interface. This consideration favors the adaptation of the whole interactive system and is reflected in the last step in the presentation model.

Thus, we found that some reviewed approaches have proposed mechanisms of learning [11] to improve the predefined knowledge to respond to the needs of users who evolve over time. In addition, the meta-UI dimension represented in the literature [1] determines the level of control available to the user on the system. In fact, the knowledge base and Meta-UI can improve the predefined knowledge by learning the habits and preferences of users. However, most approaches have not taken these two dimensions in their work.

IV. CONCLUSION AND FUTURE WORK

In this article, we studied several approaches and classified the adaptation approach based on different criteria. The criteria are used to extract the shortcomings of the interactive systems adaptation approaches proposed in literature.

In conclusion, the adaptation is not limited to the UI component of an interactive system. The adaptation to the context of use may also impact the functional core. In the light of the criteria mentioned above, it is important to note that although there are many works in the field of research, there is no approach that is based on the model to our knowledge. There is a lack of an approach which adapts the interface and functionality of the interactive system at run time. So, in the future researches we will try to propose an approach that responds to these needs.

REFERENCES

[1] L. Balme, A. Demeure, N. Barralon, J. Coutaz, and J. Calvary, “CAMELEON-RT : a software architecture reference model for distributed, Migratable, and Plastic User Interfaces”, Second European symposium on ambient intelligence, EUSAI 2004, Eindhoven, the Netherlands, November 8-11, 2004, pp. 291-302.

[2] W. Bouchelligua, A. Mahfouthi, and L. Benammar, “An MDE approach for user interface adaptation to the context of use”, third international conference, HCSE 2010, Reykjavik, Iceland, October 14-15, 2010, pp. 62-78.

[3] G. Calvary, J. Coutaz, D. Thevenin, Q. Limbourg, L. Bouillon, and J. Vanderdonckt, “A Unifying Reference Framework for Multi-Target User Interfaces”, proceeding in Elsevier journal, Interacting with Computers, vol. 15, issue 3, June 2003, pp. 289-308.

[4] J. Criado, C. V. Chicote, N. Padilla, and L. Iribarne, “A model driven approach to Graphical User Interface Runtime Adaptation”, international conference on Model Driven Engineering Languages and systems, vol, 641, ISSN 1613-0073, Oslo, Norway, October 5-13, 2010, pp. 49-59.

[5] O. Dâassi, G. Calvary, J. Coutaz, and A. Demeure, “ Towards a new generation of widgets for supporting software plsticity: the “comet””, In EHCI- DSVIS’04, the 9th IFIP working conference on engineering for human computer interaction and interactive systems, vol. 3425, Hamburg, Allemagne, October 2004, pp. 306-324.

[6] A. K. Dey, and G. D. Abowd, “Towards a better understanding of context and context-awareness”, First international Symposium, HUC’ 99, Karlsruhe, Germany, September 27-29, 1999, pp. 304- 307.

[7] J. M. Favre, “Towards a basic theory to modeldriven engineering,” In 3rd Workshop in Software Model Engineering 2004, joint event with UML2004, Lisboa, France, 2004, pp. 9-17

[8] Y. Gabillon, M. Petit, G. Calvary, and H. Fiorino, “Automated planning of user interafce composition”, International conference on intelligent user interfaces, Palo Alto, CA, Etats- Unis, February, 2011, pp. 3-8.

[9] J. V. Gurp, J. Bosch, and M. Svahnberg, “On the notion of variability in software Product Lines”, Working IEEE/IFIP Conference on Software Architecture, August 28- 31, 2001, pp. 45-54.

[10] S. Hachani, S. Dupuy-Chessa, and A. Front, “ a generic approach to HCI dynamic adaptation to context,” University of Grenoble, CNRS, LIG BP72, 38402 Saint Martin Héres Cedex, France, 2009.

[11] A. Hariri, D. Tabary, S. Lepreux, and C. Kolski, ”context aware business adaptation toward user interface adaptation,” siwn Communication, vol 3, juin 2008. pp 46-52.

[12] S. J. Pyykkö and T. Vainio, “Framing the Context of Use for Mobile HCI”, In International Journal of Mobile Human Computer Interaction, vol. 2, issue 4, October 2010, pp. 1-28.

[13] J. S. Sottet, G. Calvary, J.M. Favre, and J. Coutaz, “Megamodeling and Metamodel-Driven Engineering for Plastic User Interfaces: Mega-UI,” Human-Centered Software Engineering”, Human Computer Interaction Series, London, 2009, pp. 173-200.

[14] P. Szekely, “Retrospective and challenges for model based Interface Development”, Design, specification and

verification of interactive systems 96, Eurographics workshop in Namir, Belgium, June 5-7, 1996, pp. 1-27.

[15] D. Thevenin, J . Coutaz, and G. Calvary, " A reference framework for the development of Plastic User Interfaces", In multi-device and multi-context user interfaces: engineering and applications frameworks, Aug 2005, pp. 29-49, doi: 10.1002/0470091703.ch3.

[16] UIMS," A metamodel for the runtime architecture of an interactive system: the UIMS tool developpers workshop", ACM SIGCHI Bulletin, vol. 24, issue 1, New York, USA, Janury 1992, pp. 32-37.

[17] S. Vale and S. Hammoudi, "Context-aware Model Driven Development by Parameterized Transformation", IEEE internationnal Conference on Web Services, 2007, pp. 302-329.

[18] J. Vanderdonckt, D. Grolaux, P. Van Roy, Q. Limbourg, B. Macq, and B. Michel, "A Design Space for Context-sensitive User Interfaces", ISCA 14th Int. Conf. on Intelligent and Adaptive Systems and Software Engineering, ISCA, Toronto, 2005, pp. 207-214.

[19] M. Weiser, "The Computer for the 21st Century", ACM SIGMobile Mobile computing and communications Review, vol. 3, issue 3, New York, USA, Jully 1999, pp. 3-11.

A Tangible Directional-View Display for Interaction

Youngmin Kim, Byoungha Park, Kwang-Soon Choi, and Kwang-Mo Jung
Realistic Media Platform Research Center
Korea Electronics Technology Institute
Seoul, Korea
E-mail: rainmaker@keti.re.kr, bhpark@keti.re.kr, lenon@keti.re.kr, jungkm@keti.re.kr

***Abstract*—A tangible directional-view display system that can provide different perspective views without any special glasses is introduced. The proposed system can display perspective floating five images in the space in front of the system with the help of concave mirrors. In addition, the proposed system adopted an ultrasonic focusing technology in order to provide immersive experiences and deliver the sense of touch. We will explain our proposed method and provide theoretical analysis that supports it.**

Keywords-Interaction, Directional-View, Ultrasonicsound

I. INTRODUCTION

Since the successful development of a stereoscopic three-dimensional (3D) film, studies on autostereoscopic 3D display have been actively conducted recently. The ultimate goal of such a 3D display is to provide multiple viewers with a tangible 3D display. However, it is hard to implement such a 3D display because of several constraints, such as a lack of display panels that are usable commercially for ultimate 3D display [1-6]. Also, there are several reasons why some people are still against stereoscopic 3D displays; one of the reasons is the discomfort of wearing glasses.

For the ultimate tangible 3D display, there are several requirements. First, it is necessary to display different perspective images according to the users even though it is not necessary to provide such volumetric 3D display. Second, it should detect a hand gesture to identify the user's movements. Finally, a proper sense of touch should be provided to the viewer's fingertip in order to deliver immersive experiences [7].

In this paper, we propose a tangible directional viewable display for interaction using floating displays and ultrasound transducers. For providing high-definition directional-view images that can be viewed without any special gadgets by plural viewers, we integrated three lenticular lens displays and five floating displays into the proposed system. The proposed system has two attached infrared (IR) cameras as detectors. In addition, a number of ultrasound transducers are designed to feel tactile stimulation at a specific point. The paper is structured as follows. In Section II, we explain the proposed method by using floating displays and IR camera. In Section III, the experimental results that support the proposed method are provided, and, in Section IV, we conclude the paper.

II. PROPOSED METHOD

Figure 1 shows a schematic representation of our proposed system. The system consists of a display system in which three lenticular lens displays and five floating displays are combined, a hand gesture recognition system for interaction (IR camera), and a spatial tactile system that provides tactile expression. Various types of 3D displays, such as a holographic display and volumetric displays, can be considered to construct a tangible 3D display. However, each candidate has limitations, such as a massive amount of information to be processed and difficulty to interact with rotating screen. A floating display is a method of projecting clear two-dimensional (2D) images, which has been applied as a "Pseudo hologram" recently. Among the various abovementioned display methods, a floating display that provides vivid images and is suitable for interaction as a tangible display was selected.

Figure 1. Schematic diagram of the tangible directional-view display system

Three lenticular lens displays were adopted for delivering immersive experiences to the viewers. They located in an upper position for a number of viewers to observe images. These three lenticular lens displays were used for the selection of image contents, as shown in Figure 2. In the display system, viewers selected contents using a lenticular lens display, followed by touching the image contents directly. A floating display can be constructed by a concave mirror. This mirror is optically equivalent to the floating lens, so the location and the size of the floating image can be mathematically by the mirrors. For simplicity, a Jones transfer matrix of the floating display along the light propagation paths is expressed as follows:

$$\begin{pmatrix} x_4 \\ \theta_4 \end{pmatrix} = \begin{pmatrix} 1 & d_3 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & d_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \begin{pmatrix} 1 & d_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \theta_1 \end{pmatrix} \quad (1)$$

where x_1, x_2, x_3, and x_4 denote the location of the original image, the lens, the floating image, and the viewer's eye, respectively [8]. θ_1, θ_2, θ_3, and θ_4 denote the angle of each optical component, and r represents the radius of the concave mirror. H_1, H_2, H_3, and H_4 denote the size of the original image, the floating lens, the floating image, and the distance from the viewers' eye, respectively. According to the equation, the location and the size of the floating image can be calculated. Because the horizontal viewing angle of each floating display is 45 degree, we choose five floating displays for covering five viewers simultaneously. This floating display is varied at the lower part for viewers to see only the floating images.

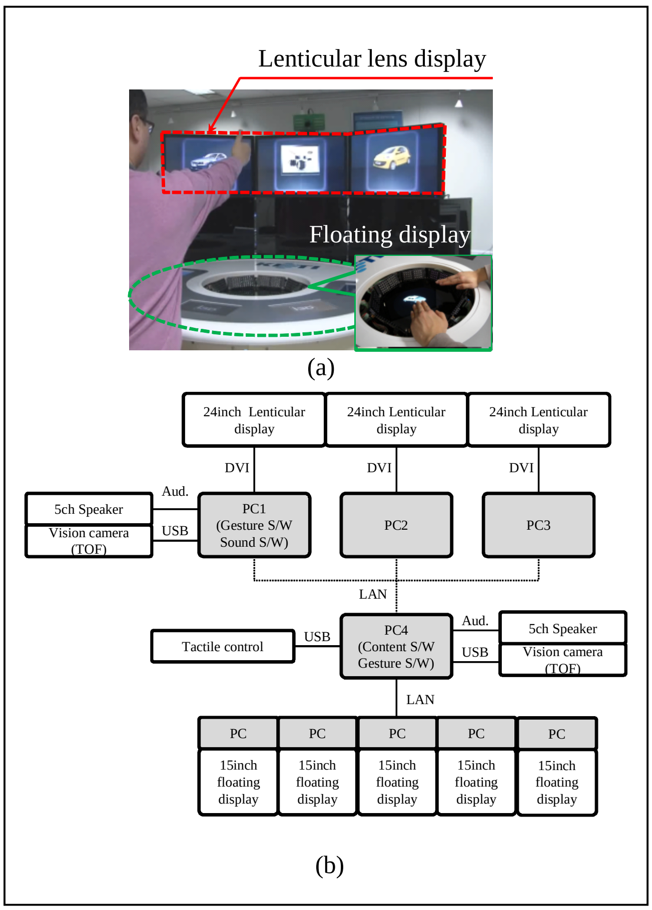


Figure 2. (a) a photo of our proposed system and (b) a system block diagram of our proposed system.

For the tactile impression, we developed a technology that can focus stimulation on a specific location using an ultrasound transducer technology [9]. The focus where a floating image is projected over a space is set up followed by obtaining a distance difference from the focus to the corresponding ultrasound transducers as shown in Figure 3. Since the distance is different according to the corresponding ultrasound transducers, a phase difference is generated per ultrasound transducer. Therefore, we should adjust this difference by using phase differences. We use 5 by 10 transducers per module and 8 ultrasound transducer modules in total were used for increasing the tactile pressure.

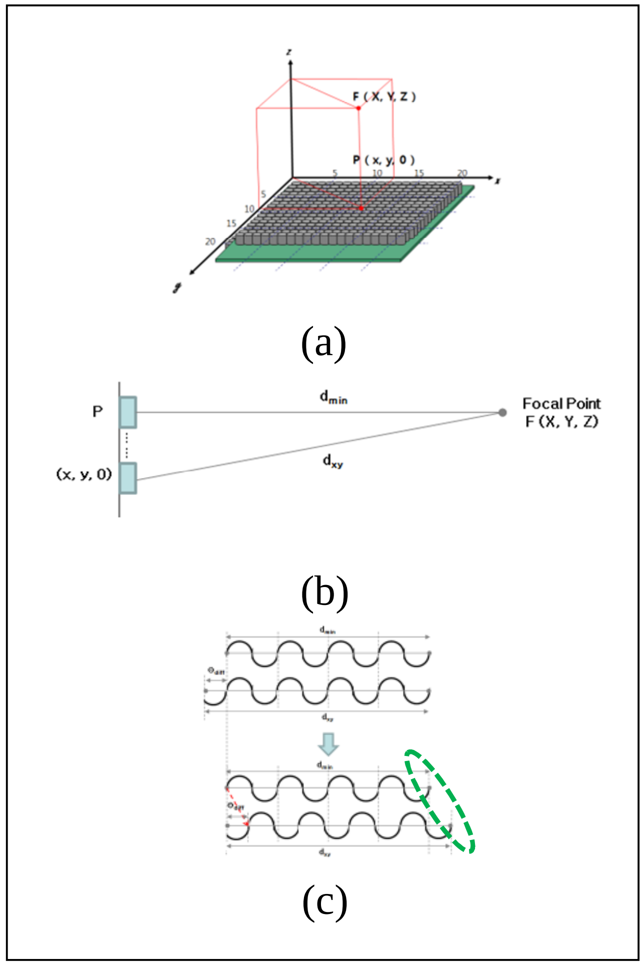


Figure 3. Focusing of ultransound wave by ultrasound transducers: (a) a schematic of ultrasound transducer system, (b) distance estimation between focal point (x, y, z) and each transducer, and (c) phase alignment of ultrasound wave by considering phase difference among transducers.

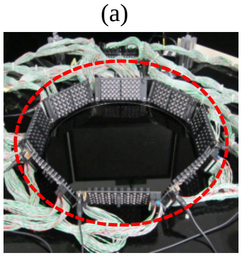
Lastly, a detecting system using IR emission units was used for applying tactile stimulation by the location of the fingertip of users. The system was designed to find out user's fingertip using general time-of-flight (TOF) mode. Four of predefined hand gesture (left, right, enter, and backward) were selected according to the user's gesture.

 ISBN: 978-1-61208-325-4

III. EXPERIMENTAL RESULTS

To verify our proposed system, we integrated three lenticular lens displays, detecting camera systems, floating displays, and tactile devices using ultrasound transducers, as shown in Figure 4. Figure 4 (a) depicts the configuration of the proposed system, and Figure 4 (b) depicts the mounted ultrasound transducers in the system. A 24 inch lenticular lens display with 9 views was used as the frontal display device. Each unit was connected with a 5-channel speaker and a TOF camera. For floating displays, 15 inch high brightness LCD monitors were used as a display panel, whose images were viewed by viewers through the polarization glass located in the center of the system. A 37 inch parabolic mirror a focal length of 9.3 inch and diameter of 35 cm in the central area was cut, and five 15 inch LCD monitors were used. Each floating display had a 45 degree field of view, and the diameter of the lower system was 90 cm, while that of the polarization glass was 30 cm. Five 10 inch tablet PCs (Samsung Galaxy note) were used and three types of scenarios were displayed by the system.

(a)



(b)

Figure 4. (a) a photo of experimental setup and (b) ten sets of ultrasound transducer module.

IV. CONCLUSION

We proposed a tangible directional-view display system for providing a user interaction experiences. The system was constructed by three sets of lenticular lens displays, five sets of floating displays, a designed ultrasound transducer system, and two TOF cameras using IR units. Integrated systems are presented to verify the validity of the proposed method, and the experimental results show that the proposed system provides different perspective view images and tactile expressions according to the predefined positions. We expect our system to have a number of applications, such as advertisement, game, e-training, and immersive digital signage industries.

ACKNOWLEDGEMENT

This research was supported by a grant from the R&D Program funded by the Ministry of Trade, Industry and Energy (MOTIE), Republic of Korea. Also, the authors are deeply thankful to all interested persons of MOTIE and NIPA (National IT Industry Promotion Agency).

REFERENCES

[1] B. Lee, “Three-dimensional displays, past and present,” Phys. Today, vol. 66, pp. 36-41, 2013, dot:10.1063/PT.3.1947.

[2] Y. Kim, et al., “A frontal projection-type three-dimensional display,” Opt. Express, vol. 20, 2012, pp.20130-20138.

[3] S. A. Benton, “Survey of holographic stereograms,” Proc. SPIE, vol. 367, 1983, pp 15-19.

[4] M. Yamaguchi, N. Ohyama, and T. Honda, “Holographic 3-D printer,” in Practical Holography IV, S. A. Benton, ed., Proc. SPIE, vol. 1212, 1990, pp. 84-92.

[5] N. Peyghambarian, S. Tay, P.-A. Blanche, R. Norwood, and M. Yamatomo, “Rewritable holographic 3D displays,” Opt. Photon. News, vol. 19, 2008, pp. 22-27.

[6] D. Lanman, M. Hirsch, Y. Kim, and R. Raskar, “Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization,” ACM Trans. Graph. vol. 29, 2010, pp. 163-172 .

[7] Y. Kim, “Tangible 3D display for interaction,” The 13th International Meeting on Information Display (IMID 2013), 2013, pp. 27-1.

[8] P. Yeh, “Extended Jones matrix method,” J. Opt. Soc. Am., vol. 72, pp. 507-513, 1982.

[9] T. Iwamoto, M. tatezono, T. Hoshi, and H. Shinoda, “Airborne ultrasound tactile display,: ACM SIGGRAPH 2008, 2008, p. 1.

HANDY: A Configurable Gesture Recognition System

Mahsa Teimourikia
Department of Electronics,
Information and Bioengineering (DEIB)
Politecnico di Milano
Email: mahsa.teimourikia@polimi.it

Hassan Saidinejad
Department of Electronics,
Information and Bioengineering (DEIB)
Politecnico di Milano
Polo Regionale di Como
Email: hassan.saidinejad@polimi.it

Sara Comai
Department of Electronics,
Information and Bioengineering (DEIB)
Politecnico di Milano
Email: comai@elet.polimi.it

***Abstract*—With the growing usage of computer systems in daily life, a natural and intuitive Human Computer Interaction (HCI) method to support the embedding of computer systems in our environment seems necessary. Gestures are of utmost importance for the design of natural user interfaces. Hand gesture recognition to extract meaningful expressions from the human hand movements and postures is being used for different applications. However, the recognition of hand gestures that contain different hand poses can be challenging. In this paper, we propose a system (called HANDY) for hand gesture recognition that is flexible to be trained to recognize a variety of user-defined gestures defined as sequences of static hand postures. The system has been designed to be used in uncontrolled environments, to handle dynamic and cluttered backgrounds, and without the need of using any wearable sensor or any specific clothing. Evaluation results show a good average accuracy in gesture recognition.**

Keywords-interactive systems; gesture-based interface; natural HCI; personalizable system

I. Introduction

Natural User Interfaces (NUI) have attracted considerable amount of research nowadays. With the growing usage of computer systems in everyday life and the desire to embed them into the environment, employing the traditional input devices (like keyboard and mouse) seems to be a bottleneck in efficient and intuitive Human Computer Interaction. Hand gestures are a natural way to communicate in human-human interactions and can be adopted also in Human Computer Interaction (HCI). More recently, devices such as Microsoft Kinect [1], a low-cost Natural Interaction (NI) device with Red-Green-Blue (RGB) and Depth sensors, have been used increasingly in NUI and HCI research for hand gesture recognition. The most prominent gesture recognition frameworks either use wearable sensors, vision-based methods, or a combination of both. While wearable sensors can be costly, inaccessible for many users and also might cause health problems or fatigue, vision-based frameworks, like ToF (Time-of-Flight) cameras [2] or the Kinect, are instead affordable to the majority of the people.

Many frameworks have been proposed regarding hand pose and gesture recognition using vision-based techniques. However, the research on gesture recognition has focused mainly on static gestures that include a static hand pose estimation or on dynamic gestures that consider the trajectory of the hands [3], [4], [5]. Postures of the hand during dynamic gesturing can be important features in natural communication. However, only a small number of frameworks have successfully taken into consideration the hand postures in the recognition of the dynamic gestures [6], [7]. And still, a limited number of hand poses are understandable for the gesture recognition system in these works. Moreover, the posture is usually assumed to remain the same during the gesturing. Even in sign language recognition systems, the posture of the hands is usually neglected and other features such as velocity, direction, trajectory, and position of the hand with respect to the face are considered. Furthermore, the flexibility of the system to be configured with user-specific gestures is yet another key aspect that has received less attention in the recent literature.

In this paper, a system (called HANDY) for gesture recognition is proposed. The framework introduced here is flexible enough to be able to train and use the system with minimal effort for a variety of hand poses and gestures that meet the user's specific needs. Microsoft Kinect depth sensor is adopted because of its low cost, availability and ease of installation. Furthermore, by the usage of depth data, the system can be adopted in low illumination, dynamic backgrounds and in uncontrolled environments and also the user would not need to use any wearable devices, sensors or any specific clothing. The mentioned aspects make the system appropriate for usage in different contexts for natural interactions.

Due to the stochastic nature of human including immeasurable and hidden mental states along with the measurable and observable human actions, Hidden Markov Models (HMM) [8] can be used to model these processes. HMMs have been successfully used in many applications and studies for speech recognition [8] and hand writing recognition [9]. Also, they have been proven effective in sign language recognition and other complex hand gesture recognition processes [10], [11]. In this approach, HMMs are used for gesture recognition to enable recognition of gestures composed of sequences of states with tolerance to the changes in the hand posture, to exploit patterns, and to evolve recognition capabilities with the help of learning techniques.

The paper is organized as follows: Section II reports related work for hand pose and gesture recognition techniques. Then, in Section III, we introduce the HANDY system and its implementation details. Section IV focuses on the results. Finally, in Section V, we draw our conclusions and report future works.

 ISBN: 978-1-61208-325-4

II. Background and Related work

In this section, we introduce the background and review the state of the art related to the main techniques needed to identify the hand and its gestures: hand localization, segmentation, hand pose estimation and gesture recognition.

Approaches regarding the interpretation of gestures in HCI either use wearable sensors or employ vision-based methods [4]. An example of wearable sensors is represented by data gloves, which can provide accurate measurements of hand pose and movements. However, wearable sensors are commonly costly while they limit the hand movement and are not very comfortable to be used in everyday life. Mynatt et al. [12] use a wireless wearable device, called Gesture Pendant, that uses infrared illumination and a charge-coupled device (CCD) camera to recognize a simple set of gestures. P. Trindade et al. [13] propose using an Inertial Measurement Unit (IMU) sensor for recognizing the hand orientation. Some other approaches employ, colored hand gloves [14], data gloves [15], and forearm band or wrist band with accelerometer and electromyography (EMG) sensors [16] accompanied with a vision-based approach to simplify the hand localization, segmentation and tracking.

To avoid the use of any wearable or mechanical device, vision-based gesture recognition can be used. Vision-based gesture recognition techniques can be divided into two broad categories [17]: Model based approaches that focus on taking advantage of a 3D or 2D model of the hand for hand pose and approaches based on hand shape appearance which is used to extract the features of the visual data for gesture recognition. Model-based approaches usually suffer from high complexity in implementation and cannot be used in live applications. Appearance-based approaches use RGB or depth data or both as the input. Our framework falls into this category.

When only RGB images in the appearance-based approach are used, the tasks of locating and segmenting the hand from a cluttered and noisy background can become very challenging, especially when other skin-colored objects are present in the scene and in case of occlusions [18]. Moreover, RGB images can be sensitive to illumination changes and extracting the sophisticated features from the image to have an accurate recognition can be very costly regarding the processing time [4]. To overcome these issues, other approaches have shown that better results for hand segmentation can be obtained when using depth data [19]. However, as it is stated in [19], most of the proposals assume that the hands are the closest object to the scene. In our system, as we will see in Section III, we will take advantage of the body tracking information extractable from Kinect SDKs, so that depth thresholding can be done regardless of the position of the hand.

Also, Zhu and Pun [20] use Kinect depth data for extracting the trajectory data sequence of the hand movements, but, do not consider hand postures. In a similar way, [21] and [7] introduce a gesture recognition approach that considers the motion and shape information of the hand using depth data. However, the hand shape is considered to remain the same during the gesturing and pose estimation is done once at the beginning of the gesture. In another study by Chen et al. [6] HMM continuous gesture recognition is proposed considering the spatial and temporal features of the gestures. Yet, a small number of poses is recognized in this approach and the posture of the hand does not change while performing the gesture. HMM has been adopted also in the work by Starner et al. [22] for American Sign Language (ASL) recognition. For recognizing the sign language they have ignored the detailed shape and pose of the hand and have only considered, coarse hand pose, orientation, and the trajectory of the gesture through time, and used such information as input for the recognition system. In a more recent work, Molina et al. [23] propose an approach for static pose estimation and dynamic gesture recognition. They successfully recognize the gestures that include the change in the hand postures. They have obtained an accuracy of 90% for recognition of combination of static hand postures and dynamic gestures.

Differently from these approaches, we consider hand poses and gestures composed of their combinations: hand poses are modeled by their skeleton, and time series analysis algorithms and HMM techniques are used to recognize sequences of hand poses. Using these techniques user-defined gestures can also be defined and more flexibility can be added in case of gesture evolution.

III. The HANDY System

The architecture of the HANDY framework for gesture recognition is depicted in Fig. 1. The system extracts from the

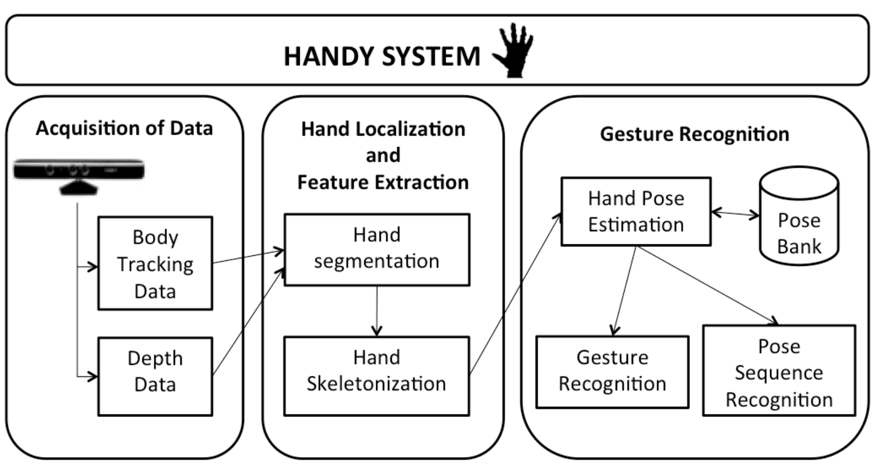


Figure 1: The Gesture Recognition Scheme

Kinect the depth data and the position of the center of the palm: the OpenNI [24] framework together with NiTE [25] skeletal tracking middleware have been used to extract the needed information. Then, the segmented hand is first identified and its skeleton is calculated as a feature that will be used for hand pose estimation. Hand poses can be stored in a pose bank, to be used in gesture recognition. Both static gestures and pose sequences are considered in our system: in both cases their recognition is based on the hand pose estimation. In the sequel, each step of this scheme is explained in detail.

A. Hand Localization and Segmentation

One of the essential components of hand gesture recognition is hand localization and tracking. Hand localization refers to the positioning of the hand and to its segmentation from the background. In other words, it means to understand which pixels of the input image belong to the hand. Using the NiTE skeletal tracking it is possible to acquire the position and depth information of the center of the palm. We employ this information to segment the hand from the background using depth thresholding. Depth thresholding is an easy and quick way for real-time hand segmentation and it can be greatly beneficial for separating the hand from the background to exclude the effects of cluttered or dynamic backgrounds. Unlike other approaches that need the hand to be the closest object to the camera or that need the person to keep his/her hand in a predefined position, here the depth thresholding is done automatically by knowing the depth information of the palm center in each frame. And therefore, no restrictions are applied for positioning the hands correctly.

Fig. 2a shows the depth data of the user that is taken from Kinect depth sensor and hand localization using the NiTE skeletal tracking. Different levels of depth are shown with different shades of gray to better show how depth thresholding can be used to segment the hand. The obtained segmented hand is shown in the upper part of Fig. 2b.

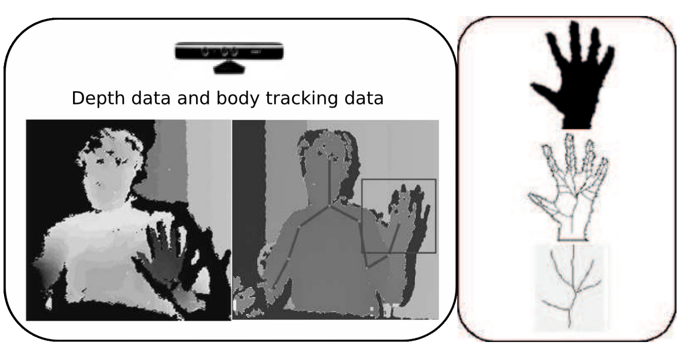
B. Feature Extraction

After hand segmentation, the system needs to extract some hand features to define and then compare different postures. In the proposed approach, skeletons are considered as the features to be used in pose estimation. A skeleton is a graph that summarizes the shape of an object and can be employed as an efficient shape descriptor for object recognition and image analysis. To obtain the skeletons of the hand (skeletonization), the Voronoi Diagram algorithm [26] is applied on the boundary points of the segmented hand. Voronoi diagrams (VD) partition the space in a number of regions based on a set of geometrical points (called seeds), in such a way that each region consists of the points closer to its seed. By applying the VD technique to the boundary points of the segmented hand, a diagram including also the main skeleton of the hand is obtained as shown in the middle part of Fig. 2b, which can be cut out by pruning the extra branches. An example of hand skeleton is shown in the bottom part of Fig. 2b.

Using this approach, defining new postures can be as easy as saving a snapshot of the posture into the system. In the HANDY framework, the user can perform the custom poses in front of the Kinect and the system saves the skeleton data of the specified poses for later use in hand pose and gesture recognition. In Fig. 3, it is possible to view some samples of hand poses that can be defined with our framework and their extracted features (skeletons).

C. Hand Pose Estimation

Hand pose estimation refers to the recognition of a single posture of the hand. This process uses the data provided by the feature extraction module, and the output will be the recognized hand pose for the performed posture. In the proposed approach, it is possible to define a set of hand poses by saving the extracted skeleton of the specific posture in a bank of hand poses. After defining hand poses, each of the performed pose skeletons will be compared with all the poses available in this bank and the most similar one will be chosen as the estimated hand pose. Since the skeletons of the hands can be slightly different even for the same hand posture performed by the same user, a method for comparing the skeleton graphs should be adopted. In our approach, to estimate the similarity of two skeletons, we adopt the Dynamic Time



(a) (b)

Figure 2: (a) Depth data and body tracking data, (b) Hand segmentation (top), application of VD on boundary points (middle) and the resulted skeleton via pruning (button)

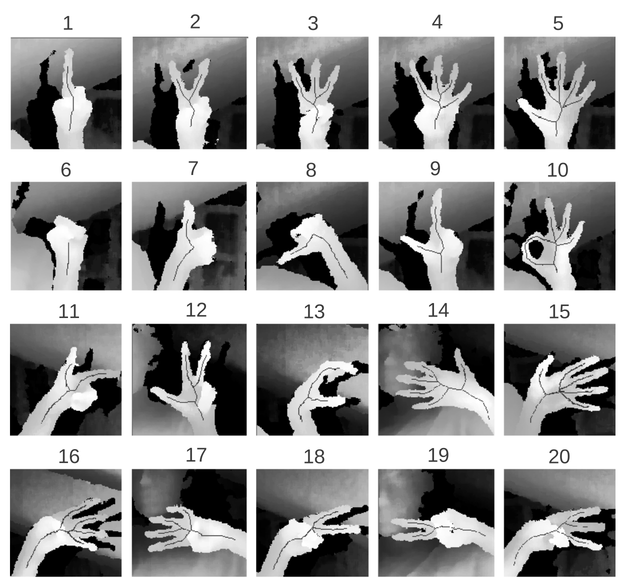


Figure 3: Some sample hand poses defined in the HANDY framework

Warping (DTW) algorithm [27], which allows to compare time series. Our algorithm takes into account the internal order of the points in the skeleton and, for a fair comparison, it normalizes the skeletons to make the the hand pose estimation system invariant to the distance of the hand to the depth sensor and also to the absolute position of the hand in the camera field of view. For further details about the process adopted in our framework the reader may refer to [28].

D. Gesture Recognition

As mentioned before, the goal of the HANDY system is to recognize both static gestures consisting of a single pose (e.g., the poses depicted in Fig. 3) or generic gestures composed of sequences of different hand postures. Some examples of these kinds of gestures are shown in Fig. 4. In both cases, it seems convenient to employ the results of the hand pose estimation as the input of the gesture recognition system.

The hand pose estimator based on DTW is able to recognize the saved postures with a reliable accuracy; however, if we see a gesture as the composition of single poses and for each sub-pose we try to recognize it by comparing it with the bank poses, there may be cases in which the hand pose does not exist in the bank (e.g., for intermediate poses); moreover, when the hand is moving between the poses, the ability of the pose estimator degrades and might not be able to recognize the correct pose. In such cases, the most similar hand posture (skeleton) will be assigned to the performed pose. Furthermore, as already noticed, there are variations in features even when the same gestures are being performed by the same user. In these cases, machine learning methods such as HMMs can be applied to handle the variations and also to recognize the gestures in which the pose estimator is unable to estimate the correct pose sequences. It is worth noting that misrecognition of an intermediate hand pose and assignment of the closest hand pose from the bank is tolerated by the gesture recognition mechanism thanks to HMMs, which are based on stochastic processes. Indeed, the HMM is a collection of states connected by transitions, that can be used to represent the statistical behavior of observable symbol sequences in terms of network states. Each observed symbol (an estimated hand pose) happens corresponding to its probability function in an HMM state. Once a symbol is observed, the HMM can stay in the same state or move to another state due to transition probabilities associated with each state [6].

Using HMMs for gesture recognition, training is needed before the system is able to recognize the gestures. As we will see in the evaluation section, a little amount of training data (in our tests, 20 for each gesture) is needed for the system to obtain an acceptable accuracy in gesture recognition. We assume that the user needs to wave, for the system to start looking for meaningful gestures. While user's hands are in the rest mode (e.g., hands are on the arms of the armchair) the gesture recognizer also goes into the rest mode. The gesture duration is considered to be fixed and we sample the data every 100 milliseconds up to 1 second and the hand pose for each acquired frame will be estimated. In this way, we will have a sequence of 10 estimated hand poses which will be used as the input for training the HMMs and for gesture recognition. The duration of the gesture and the interval of the sampling can be increased and decreased when convenient.

Fig. 5 shows a snapshot of the application that has been developed for testing the introduced system with different

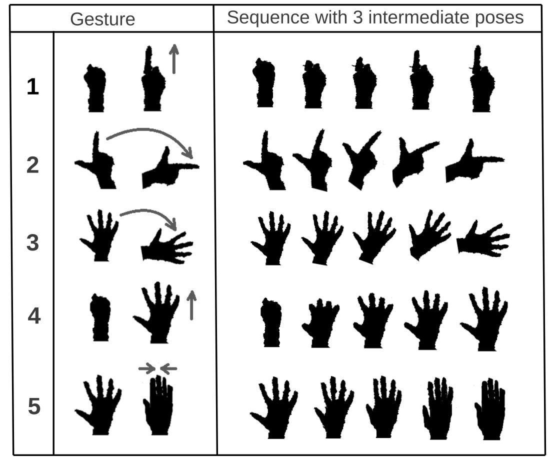


Figure 4: Set of gestures for test

users. In the application, it is possible to define and train different hand poses and gestures for any user. In the next section, the results of testing the system are discussed in more detail.

IV. Results and Discussion

For the evaluation of the system, 7 subjects were asked to train and test the system with a set of predefined gestures. All the subjects were right-handed but two of them used their left hand (thus, introducing also a lower self-confidence and fidelity in the performance of the gestures). A set of 20 hand postures, shown in Fig. 3, were recorded for each user in the pose bank. In a first experiment, the recognition of the single isolated poses was tested. In this case, the average accuracy of the pose estimation system for the sample poses in Fig. 3 was around 83.52%. The poses that include the rotation of the wrist showed less accuracy since the pose estimation system is not rotation invariant and the mentioned poses can vary regarding rotation even when performed by the same user. The gesture recognition is tested according to the before mentioned user specific data. However, as we will see next, when a gesture defined as a sequence of poses is considered and during the gesture intermediate hand poses are sampled with a fixed interval, the accuracy can be increased.

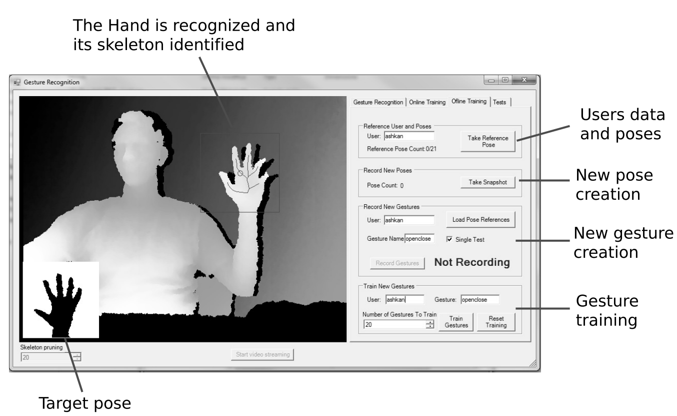


Figure 5: The application for testing the HANDY gesture recognition system

The poses to be stored in the system should be, however, defined according to the nature of the gestures. The more poses are defined in the system, the more similar they get to the postures that are used in gestures, the more accurate will be the inputs of the gesture recognition system. However, there is a trade-off between the accuracy of the input pose sequence and the number of the poses that are defined, as with a large pose bank the pose estimation system can get slower.

In our tests, we have used the 20 poses depicted in Fig. 3 and, based on these poses, tried to recognize the five gestures shown in Fig. 4. These five gestures were chosen taking into account that:

1) Poses with similar skeletons exist in the gesture. Having similar poses can confuse the pose estimation and reduce the accuracy in producing the gesture recognition inputs. For example, for the first gesture in Fig. 4, it is possible to see that its beginning and final poses (pose no. 6 and pose no. 1 in Fig. 3) have very similar skeletons that are composed of a single vertical branch.
2) Some of the poses in the gesture might not exist in the hand pose bank used for pose estimation. For example, the gesture number 5 in Fig. 4 includes poses like the closed hand, the final pose in the sequence, which is not included in the gesture set recorded in the pose bank (depicted in Fig. 3). If a pose does not exist in the pose bank, the most similar pose from the bank will be selected in the pose estimation step.
3) Gestures contain some intermediate poses with lower pose estimation accuracy. As an example, in the gesture number 3 depicted in Fig. 4 the final hand pose (that corresponds to pose number 15 in Fig. 3) has shown a low accuracy of only 40% in the pose estimation system. This is also true for the second gesture in Fig. 4.

These gestures were recorded by a single user as isolated gestures and then used for training and testing the system. In the test, gestures have a fixed length of 1 second, and are sampled 10 times during their performance. In a real application, the number of of sampling and the duration of the gestures can be configured in the system according to the nature of the gestures and the user specific considerations. Evaluation of the HANDY was done in two steps: first, the system is trained for a single user using 20 training data for each gesture. Later, the system is tested for the same user.

Table I shows the accuracy of each gesture per user: the first column represents the user; the second indicates if the left or right hand was used; the next columns represent the five gestures specified in Fig. 4: for each gesture the percentage indicates the accuracy in the recognition of 30 performances of that gesture.

Table II demonstrates the average accuracy of the same results for the left and right-hand users: from the results, no considerable differences can be concluded in the accuracy of the system in the two cases.

Finally, Table III reports the average confusion matrix for the tested gestures with the different users: for example, gesture G1 was recognized correctly in the 94.28% of the cases, it was confused 3.33% with gesture G2 and 2,86% with gesture G4. The average accuracy of 95.57% was concluded during these tests which is acceptable for many applications.

In this first evaluation for each user, we considered the hand poses and the gestures previously defined and trained by the user him/herself. In a second evaluation, we tried to evaluate the accuracy of gesture recognition considering the case of a new user who has not trained the system before. For this evaluation the system was trained with the data from all the subjects except the user who is going to test the system. Only the data from right hands were used since we did not have enough training data for left hands. The results of this evaluation are reported in Table IV. The average recognition accuracy of 96.26% was obtained. This suggests that with more training data some universal gestures could be recognized without initial training.

While testing the system the users reported that they did not feel any delay in the gesture recognition system and therefore that the system is able to respond in real-time. Moreover, subjects were asked to rank the gestures based on their difficulty. Those involving difficult hand manoeuvres like wrist rotation (gestures 2 and 3, see Fig. 4) were commented to be the most difficult ones. However, in general, they found the system easy to use.

V. Conclusions and Future Works

In this work we have presented an approach for a flexible and configurable hand pose and gesture recognition system that

TABLE I: RECOGNITION ACCURACY OF TRAINED GESTURE RECOGNITION SYSTEM FOR SINGLE USERS

Users	R/L-hand	G1	G2	G3	G4	G5
U1	R	100%	90%	100%	96.67%	100%
U2	R	100%	100%	100%	100%	100%
U3	R	100%	100%	96.67%	96.67%	100%
U4	R	90%	100%	100%	100%	100%
U5	R	83.33%	100%	100%	76.67%	96.67%
U6	L	86.67%	80%	100%	100%	100%
U7	L	96.67%	100%	100%	96.67%	93.33%

TABLE II: AVERAGE RECOGNITION ACCURACY OF TRAINED GESTURE RECOGNITION SYSTEM FOR SINGLE RIGHT-HAND AND LEFT-HAND USERS

R/L-hand	G1	G2	G3	G4	G5
R	94.66%	99.95%	99.33%	94%	93.33%
L	91.67%	90%	100%	98.33%	96.66%

TABLE III: AVERAGE CONFUSION MATRIX OF GESTURE RECOGNITION SYSTEM FOR SINGLE USERS

Gestures	G1	G2	G3	G4	G5
G1	94.28%	3.33%	0%	2.86%	0%
G2	0%	95.71%	0.48%	0.48%	2.86%
G3	0%	0%	99.59%	0%	0.48%
G4	0%	0.95%	0%	95.24%	3.81%
G5	0%	0.48%	0.48%	0.48%	98.57%

TABLE IV: RECOGNITION ACCURACY OF TRAINED GESTURE RECOGNITION SYSTEM FOR SINGLE USERS

Users	G1	G2	G3	G4	G5
U1	100%	90%	100%	100%	93.33%
U2	100%	100%	100%	96.67%	100%
U3	100%	100%	100%	86.67%	96.67%
U4	100%	96.67%	93.33%	80%	90%
U5	96.67%	93.33%	93.33%	100%	100%

could be adapted to the specific needs of the user. Specific hand postures can be defined and recognized in this system according to the needs of the user, the application and nature of the dynamic gestures that are based on these postures. In our tests, 20 different hand poses have been considered: the comparison of the performed pose with the skeleton stored in the pose bank led to accurate results in 83.52% of the cases. In addition, the system can be trained with dynamic gestures, considered mainly as a continuous sequence of hand poses, with small number of training gestures (20 to 30) to achieve a good recognition accuracy. We obtained a good accuracy of 95.57% in our tests, involving even gestures with intermediate poses having lower estimation accuracy.

As future work, we plan to improve the hand pose estimation to be able to also recognize the postures in which the fingers might not be visible. The gesture recognition considered in this paper can be extended to consider also hand trajectories, to be able to recognize a larger variety of gestures. We have in mind also to do a comparative evaluation of our recognition system with the similar ones. Further extensions of this system may include also facial expression recognition and vocal orders to support an even wider set of applications that use natural user interfaces.

References

[1] "Kinect for windows, official website, http://www.microsoft.com/en-us/kinectforwindows/," accessed: 2014.01.28.

[2] S. B. Gokturk, H. Yalcin, and C. Bamji, "A time-of-flight depth sensor-system description, issues and solutions," in Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on. IEEE, 2004, pp. 35–35.

[3] S. Mitra and T. Acharya, "Gesture recognition: A survey," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 37, no. 3, 2007, pp. 311–324.

[4] G. Murthy and R. Jadon, "A review of vision based hand gestures recognition," International Journal of Information Technology and Knowledge Management, vol. 2, no. 2, 2009, pp. 405–410.

[5] Y. Wu and T. S. Huang, "Vision-based gesture recognition: A review," Urbana, vol. 51, 1999, pp. 103–115.

[6] F.-S. Chen, C.-M. Fu, and C.-L. Huang, "Hand gesture recognition using a real-time tracking method and hidden markov models," Image and Vision Computing, vol. 21, no. 8, 2003, pp. 745–758.

[7] N. D. Binh, E. Shuichi, and T. Ejima, "Real-time hand tracking and gesture recognition system," Proc. GVIP, 2005, pp. 19–21.

[8] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," Proceedings of the IEEE, vol. 77, no. 2, 1989, pp. 257–286.

[9] A. Kundu, Y. He, and P. Bahl, "Recognition of handwritten word: first and second order hidden markov model based approach," Pattern recognition, vol. 22, no. 3, 1989, pp. 283–297.

[10] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden markov model," in Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on. IEEE, 1992, pp. 379–385.

[11] J. Schlenzig, E. Hunter, and R. Jain, "Recursive identification of gesture inputs using hidden markov models," in Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on. IEEE, 1994, pp. 187–194.

[12] E. D. Mynatt, A.-S. Melenhorst, A.-D. Fisk, and W. A. Rogers, "Aware technologies for aging in place: understanding user needs and attitudes," Pervasive Computing, IEEE, vol. 3, no. 2, 2004, pp. 36–41.

[13] P. Trindade, J. Lobo, and J. P. Barreto, "Hand gesture recognition using color and depth images enhanced with hand angular pose data," in Multisensor Fusion and Integration for Intelligent Systems (MFI), 2012 IEEE Conference on. IEEE, 2012, pp. 71–76.

[14] C. Keskin, A. Erkan, and L. Akarun, "Real time hand tracking and 3d gesture recognition for interactive interfaces using hmm," ICANN/ICONIPP, vol. 2003, 2003, pp. 26–29.

[15] F. Camastra and D. De Felice, "Lvq-based hand gesture recognition using a data glove," in Neural Nets and Surroundings. Springer, 2013, pp. 159–168.

[16] Y. Li, X. Chen, J. Tian, X. Zhang, K. Wang, and J. Yang, "Automatic recognition of sign language subwords based on portable accelerometer and emg sensors," in International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, no. 17. ACM, 2010.

[17] P. Garg, N. Aggarwal, and S. Sofat, "Vision based hand gesture recognition," World Academy of Science, Engineering and Technology, vol. 49, no. 1, 2009, pp. 972–977.

[18] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," Computer Vision and Image Understanding, vol. 108, no. 1, 2007, pp. 52–73.

[19] J. Suarez and R. R. Murphy, "Hand gesture recognition with depth images: A review," in RO-MAN, 2012 IEEE. IEEE, 2012, pp. 411–417.

[20] H.-M. Zhu and C.-M. Pun, "Real-time hand gesture recognition from depth image sequences," in Computer Graphics, Imaging and Visualization (CGIV), 2012 Ninth International Conference on. IEEE, 2012, pp. 49–52.

[21] X. Liu and K. Fujimura, "Hand gesture recognition using depth data," in Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on. IEEE, 2004, pp. 529–534.

[22] T. Starner, J. Weaver, and A. Pentland, "Real-time american sign language recognition using desk and wearable computer based video," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 20, no. 12, 1998, pp. 1371–1375.

[23] J. e. a. Molina, "Real-time user independent hand gesture recognition from time-of-flight camera video using static and dynamic models," Machine Vision and Applications, vol. 24, no. 1, 2013, pp. 187–204.

[24] "Openni official website, http://www.openni.org/," accessed: 2014.01.28.

[25] "Nite official website, http://www.openni.org/files/nite/," accessed: 2014.01.28.

[26] F. Aurenhammer, "Voronoi diagrams: A survey of a fundamental geometric data structure," ACM Computing Surveys (CSUR), vol. 23, no. 3, 1991, pp. 345–405.

[27] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series." in KDD workshop, vol. 10, no. 16. Seattle, WA, 1994, pp. 359–370.

[28] M. Teimourikia, "Design and development of a kinect-based multi purpose gesture interface for fragile people," Master's thesis, Politecnico di Milano, October 2013.

 ISBN: 978-1-61208-325-4

Concepts of Multi-artifact Systems in Artifact Ecologies

Henrik Sørensen and Jesper Kjeldskov
Research Centre for Socio-Interactive Design
Department of Computer Science
Aalborg University, Denmark
{hesor, jesper}@cs.aau.dk

***Abstract*— The artifact ecologies emerging from the increasing number of interactive digital artifacts, capable of communicating with each other, have created a situation where software applications no longer need to be limited by the physical boundaries of a single artifact. In order to take advantage of the full potential of this situation, we first need to establish a common understanding of the interaction that crosses physical artifact boundaries. Eventually, this will help us understand and design multi-artifact systems that are more than the sum of its individual parts. In this paper, we analyze two multi-artifact systems from our prior work within the domain of music consumption and identify four concepts of multi-artifact interaction: Plasticity, migration, complementarity, and multi-user. We discuss the concepts in order to relate them to an artifact ecology thinking and identify implications for future work.**

Keywords- artifact ecology; multi-artifact; music system

I. INTRODUCTION

The establishment of a wireless network infrastructure surrounding us has introduced easier connectivity between different digital devices. In addition, to enable data sharing and synchronization it provides great potential for interactions transcending the physical boundaries of individual devices. Jung et al. [8] describe this network of devices as a personal ecology of interactive artifacts and defines it as *"a set of all physical artifacts with some level of interactivity enabled by digital technology that a person owns, has access to, and uses"*. Taking advantage of the potential offered by such artifact ecologies is however challenging. Our focus lies in the concepts of the interaction between humans and artifacts. It is however clear that interaction designs spanning multiple artifacts is highly dependent on a comprehensive and flexible technical infrastructure for artifact discovery, connection, and communication. We therefore work under the assumption that this is or will be available to some extent.

Interaction designers have become quite good at designing desktop applications and are in a post-desktop era progressively getting better at designing mobile artifacts as well. It is however, our belief that good interaction design for artifact ecologies consists of more than the aggregation of good designs for each individual artifact. Previous work has already moved towards an understanding of the composition [8] and dynamics [4] of the ecologies as a whole. What we find is however that there is a gap between the work on understanding interactions with single artifacts and understanding our personal artifact ecologies. Understanding multi-artifact systems that combine specific artifacts from our personal artifact ecologies creates an additional layer of complexity that requires us to think of the system in a holistic way on an abstraction layer above the single artifact but below the entire artifact ecology.

The overall goal is to move towards multi-artifact interaction designs that deliberately exploit the synergetic effects of artifact combinations. Our contribution in this paper is a step towards an understanding that eventually can lead to this goal. The specific contribution is to identify and discuss concepts of multi-artifact systems that we find to be of particular significance to an artifact ecology context. We base our analysis around multi-artifact systems from our previous work in the music domain.

First, we present related work on artifact ecologies and music consumption in Human-Computer Interaction (HCI). We then clarify our understanding and delimitation of the multi-artifact system concept followed by a description of the two music systems from our prior work. Finally, we analyze the systems in order to identify and discuss characteristic concepts of multi-artifact systems before we conclude with implications for future work.

II. RELATED WORK

This section relates our work to previous research in artifact ecologies and music consumption.

A. Artifact Ecologies

In a study of the social role of products, Forlizzi [6] introduced a product ecology framework used to describe the dynamic aspects of use. The framework puts the product in the middle, meaning that each individual product has its own ecology in which components are interconnected. For example, a product often has certain relations to other products that together act as a system. The components included in the framework, besides other products, are people, activities, place, and the routines and cultural context. Forlizzi's product ecology framework provides means to reason about the single product and its social impact across users.

Artifact ecologies represent a different approach of putting an ecological thinking into play in relation to the

 ISBN: 978-1-61208-325-4

products surrounding us. In the framework of Jung et al. [8], they put the user in the center and define a personal ecology of interactive artifacts that a person owns, has access to, and uses. This means that an ecology is defined from the perspective of a person instead of a product/artifact. In their work, they conducted two types of exploratory studies with the common goal of understanding the relationships within artifact ecologies. Their study works under the assumption that the experience with an artifact can only be fully understood when it is considered in relation to an artifact ecology. We find the personal perspective very useful in understanding interactions that involve several artifacts. However, the limitations of the framework are that it does not take into account what happens when different personal ecologies intersect in multi-user interactions.

Jung et al. [8] argues that artifact ecologies are dynamically evolving. Bødker and Klokmose [4] follow up on that idea and emphasize the importance of not only understanding a current composition of artifacts in our surroundings but also how relationships among them change over time. Using Activity Theory as their theoretical framing and the Human-Artifact Model [3] as an analytical tool, they identify three states of an artifact ecology: The *unsatisfactory*, the *excited*, and the *stable* state. The artifact ecology of a person will change state over time and at some point reach the unsatisfactory state once again. Changes to the ecology can then put it into an excited state and the cycle repeats itself. One challenge they encountered in their analysis was to describe what the artifacts of artifact ecologies is. While Jung et al. [8] describes artifacts as physical objects, Bødker and Klokmose [4] found from their study that this did not tell the whole story and that something more may be needed to systematically address the software as well.

B. Music Consumption

Music has always been an interesting topic due to its universal appeal. Holmquist talks about the field of ubiquitous music and how it has been formed by a digitization of music, portable music players and heightened bandwidth [7]. Although the article is from 2005, the notion of ubiquitous music has only become more relevant due to the emergence of Internet streaming services and affordable multi-room music systems. Liikkanen [10] however points out that music consumption, as a defined research area in HCI is extinct. He argues that research on music consumption through interactive devices continues but is marginal. There are however still interesting projects emerging in the HCI community. An example that operates in the area of multi-artifact interactions is Mo by Lenz et al. [9]. Mo is a music player with an integrated speaker that focuses on a shared music experience. Mo can be brought into a social setting and by placing it next to other Mo players, it creates a connected music system.

III. Multi-Artifact Systems

Before we start conceptualizing multi-artifact systems in artifact ecologies, it is important for us to clarify what we mean by the term in the first place and how we delimit it to reflect our scope. By multi-artifact systems, we refer to interactive systems, which are part of an artifact ecology, and involves more than a single physical artifact. Different terms have previously been used to describe similar concepts. Rekimoto has for instance described it as multiple-computer user interfaces with a focus on graphical user interfaces [11]. Terrenghi et al. [12] have furthermore created a taxonomy for what they refer to as multi-person-display ecosystems. As much as we appreciate the desirable features of the visual aspect, we also want to acknowledge other modalities of input and output, especially since our point of departure is in the music domain. Multi-device interface is another term often used. It however ambiguously describe also interfaces accessible across different platforms, which is not part of our scope. Because we want to continue the ongoing work on artifact ecologies, it makes sense to refer to the sub-sets of artifacts as multi-artifact systems. According to the systems' view, the essential properties of an organism, or a system, is the properties of the whole that none of the parts has alone [5]. This view fits perfectly well with our intention of addressing systems with interaction designs that provide more than cross-platform interfaces.

A. Delimitation

We acknowledge Bødker and Klokmose's [4] notion of the artifact term encompassing more than the physical interactive artifact. Our interest lies in the interaction designs, which transcends the boundaries of a physical artifact, thus we use multi-artifact systems as a term to describe sub-systems of artifact ecologies consisting of a specific composition of hardware *and* software artifacts used throughout a particular activity. This could technically involve the interaction with a desktop-PC communicating with a web server through a browser, but our focus is more specifically on systems where either the user provides direct input to the artifacts or the artifacts provide direct output to the user. The server part of the example fulfils neither role. Another example is video conferencing that involves several artifacts, however only one from each user's personal ecology, hence it is not a multi-device system either. A system that merges two persons' smartphones into a common interface when put together is however an example of a multi-artifact system from our perspective, as it would become a multi-artifact system in each user's ecology. The last example shows the inclusion of systems that exist in the intersection between different personal artifact ecologies, where multiple persons interact with some or all of the same artifacts.

 ISBN: 978-1-61208-325-4

B. *Time and Space*

Although the browser and video conferencing examples provide some limitation to our scope it is not to be understood as if the artifacts in the multi-device systems are required to be co-located or even that the interaction with each artifact has to happen simultaneously. We still consider systems that distribute interaction across time and space. It will however have to be as a part of the same activity from the personal point-of-view. An important point is also still that the system should provide more than an interface accessible from different artifacts. An example is the Google Chrome browser. Having a version for Windows, Android and iOS is still a single-device interaction, but when it starts remembering open tabs, bookmarks, search preferences etc. across artifacts it becomes a multi-artifact system.

The following two sections provide descriptions of the two multi-artifact music systems from prior work, on which we base our conceptualization.

IV. AIRPLAYER

AirPlayer is a multi-room music system that adapts to the location of the user with the purpose of allowing for an implicit interaction. It builds on top of Apple's AirPlay, hence the name, making it capable of streaming music from a central digital music collection to speakers placed in different locations around the home. Each speaker connects wirelessly to a central music player application through an Airport Express that also works as a Wi-Fi access point. The use of a Wi-Fi network furthermore makes it possible for the user to control the music independently in specific locations from a smartphone application. AirPlayer handles separate locations through the notion of *zones.* A zone is per default a representation of the room in which a particular Airport Express is placed. The user can however combine zones to play and control the music in several locations simultaneously. The zone name is visible in the bottom of the main screen (see Fig. 1). By sliding horizontally, the user can cycle through the different zones to see the current song playing, change the volume etc.

Similar features are already present in Apple's existing product family, through iTunes, as well as in other multi-room music systems. What is significant to AirPlayer is the addition of the proxemic interaction features that allows the system to adapt to spatial relations between the user and particular speakers placed in different rooms. The proxemic interaction manifests itself in AirPlayer as two features called *location* and *movement*, which the user enables through the smartphone application. A simple implementation of an indoor positioning system. The smartphone application continuously measures Received Signal Strength Indicator (RSSI) values from the Airport Express Wi-Fi access points and uses them to estimate in which zone the user is located.

A. *Location*

When the location feature is enabled, the smartphone application continuously adapts to represent the music currently playing in the zone where the user is located. As illustrated in Fig. 1, this means that the user interface presents information about the song playing and furthermore automatically controls the music in this particular location. The change happens in a seamless and subtle way, when the user changes location, without the need for further explicit user interaction. Whenever the system detects a change in location, it simply adapts the smartphone application to represent the current zone. The interaction from the user point-of-view is similar to having a universal remote control for independent music players in each room.

B. *Movement*

When the movement feature is enabled, the music follows the user around the home as illustrated in Fig. 2. By tracking the smartphone, the system is able to anticipate which zone the user is entering, continue the music in the new zone, and stop the music in the old one. What actually happens is that AirPlayer streams the music to all zones simultaneously and simply adjusts the volume in accordance to the location of the user. The movement and location feature can be enabled/disabled independently but are not

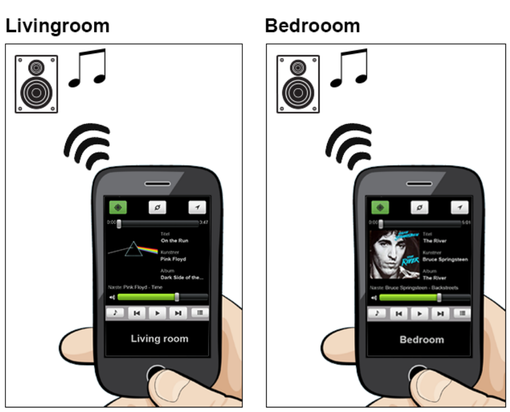


Figure 1. The location feature adapts the user interface and control to the location of the user.

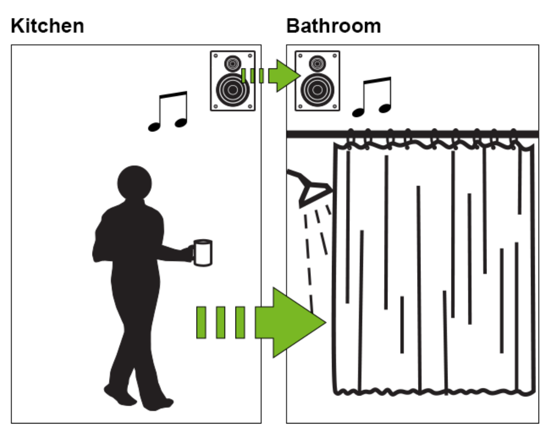


Figure 2. Music follows the user across locations.

strictly independent. Whenever the movement feature is enabled, so is the location feature as the same music is always playing where the user is located. The location feature enables a state where the smartphone user interface adapts to the location of the user and the music content stays. Inversely the movement feature enables a state where the user interface stays the same and the music content adapts to the location of the user.

V. MEET

The second system, called Music Experienced Everywhere Together (MEET), is a multi-device, multi-user music system used, to explore the interaction space of distributed interactions with co-located artifacts. The concept of MEET is to allow co-located users to share their music at social events, in order to nominate and vote for songs using their smartphones, thereby influencing the music in a collaborative manner. The interaction design consists of the following entities.

A. Smartphone Application

This smartphone application is the primary input artifact for the music player. Besides the music sharing control, it features a nominate functionality, where users can browse the collection of music shared by users and nominate songs they would like to hear. Another part of the interface present the list of nominations, giving the option to give a positive or negative vote for each nomination. Each vote will simply add or subtract one point from the total score. An important aspect is to utilize the users' own smartphones, thereby making it a personal token representing the specific owner's choices at any time.

B. Tablet Application

The tablet application is a simplified version of the smartphone application that only works for nominating and voting. It first serves the purpose of a public input artifact used by people without a compatible smartphone and secondly to create a physical interaction point for the music system in general. Because the tablet is an artifact shared among several users, we modified the vote feature to include a 10-second countdown after a vote, where the application locks itself. We added this mechanism to prevent a person from voting repeatedly for the same song.

C. Situated Display

The situated display shows the primary visual output of the music player to the users. The interface is suitable for a large flat screen TV or projector and should be placed with visibility in mind. The situated display represents nominations as album covers. The current score is represented by size, meaning that the largest nominations are more likely to be played next. This score does not map to the smartphone application, thus the situated display is the only place where the status is visible. Fig. 3 shows the voting interface of the different artifacts.

The music system is running in one place and distributes interaction to other artifacts. Specific artifacts consist of a device with a part of the distributed interface each with their own output screen and each serving a specific purpose.

VI. CONCEPTS

In this section, we use the two presented systems to identify concepts that we find meaningful in the context of multi-artifact systems. The concepts are not necessarily novel in themselves, but the contribution lies in the use of them as concepts that describe interaction across artifacts.

A. Plasticity

In AirPlayer, the location feature enables the smartphone application to adapt to the location of the user, providing control of the music in this particular location as well as information about the music playing. Balme et al. refers to this kind of adaptability as *plasticity* [1]. More precisely, they define plasticity applied to HCI as *"...the capacity of an interactive system to adapt to changes of the interactive space while preserving usability"*.

Plasticity is not only meaningful in multi-artifact systems but for single artifacts as well. A smartphone application could for instance adapt to the location of the user independently of other artifacts, or a public display could adapt to the time of day or number of people in front of it. In AirPlayer, it is the spatial relations between the smartphone and speakers placed around the home that determines what is presented to the user, which is why we argue that plasticity also has its place as a concept of multi-artifact systems.

MEET does not have any plasticity integrated in the interaction design. Each artifact has a certain form that plays a specific role in the system. An idea of introducing it into the smartphone application could however be to provide

Figure 3. The different artifacts of MEET and their respective GUIs for the voting functionality.

more feedback on the status of the voting, if the user is not able to see the situated display.

Another interesting challenge of artifact ecologies is the increase in general-purpose artifacts capable of executing different sort of applications Our phone is no longer just for making phone calls, our TV is no longer just for watching TV, and the newest addition to our ecologies is seemingly smart watches that does much more than showing the time. As our collection of general-purpose artifacts expand arguably so does the number of multi-artifact systems and the complexity of them. In AirPlayer, the smartphone application adapts to contextual information within the user's current activity. Artifacts able to adapt to fit a certain activity and composition of artifacts could be an interesting aspect of plasticity.

B. Migration

The movement feature of AirPlayer makes music follow the user around the home by moving the music output from one artifact to another. This behavior is very much in line with the work on *migratory user interfaces*, which Berti et al. describes as *"...interfaces that can transfer among different devices, and thus allow the users to continue their tasks..."* [2]. The essential issue here is the continuity in the interaction. The interesting thing about the movement feature of AirPlayer is not that it plays the same music from a central source. The interesting thing is the ability to do so continuously across locations as the user moves around. In the AirPlayer example, it is only the content (the music) that migrates and always between exactly two artifacts. Berti et al. however also defines different levels of migration:

- Total migration: The entire interface migrates from one artifact to another.
- Partial migration: Only a part migrates to the target artifact.
- Distributing migration: The interface migrates to multiple target artifacts.
- Aggregating migration: The interface migrates from multiple artifacts into one.

Migration and plasticity are somehow related concepts that both encourage more flexible and adaptable relations in our artifact ecologies. There is no implementation of interface migration in MEET but is in a similar way as plasticity a concept that could be integrated.

C. Complementarity

In MEET, the system distributes interaction across different artifacts. The different artifacts can be described as being *complementary* to each other, as each of them provides features that improve the overall system. The music player is useless if no one has connected a smartphone, shared some music and nominated at least one song. The smartphone application similarly does nothing on its own. Distributing functionality is of course a conscious design choice that is not strictly necessary to play music at a party. Doing so however takes advantage of available interaction resources to create a different kind of experience. What field studies of MEET have shown is also that such systems can provide an opportunity for a different social interaction and utilization of the environment, than a traditional music system. The benefits however come with the cost of an additional level of complexity, both technically and in the interaction design that we needs to address.

The complementarity between the smartphone/tablet and situated display in MEET is similar to the notion of *coupled displays* [12] and a lot can be learned from the work on that topic. It is however important also to consider other modalities of input and output of multi-artifact systems as artifacts may be able to utilize these to complement each other in different ways in different contexts.

AirPlayer similarly has an element of complementarity in its interaction design although more subtle than in MEET. The smartphone application provides the input and output to a music system distributed throughout the home that provides the music output. Although the smartphone application is able to control various music outputs independently, the complementarity in AirPlayer is basically a remote control metaphor. In a way this is also the case in MEET although both examples illustrate that complementary artifacts can be more powerful than a direct mapping of a traditional remote control.

It seems reasonable to talk about dependency of the relationships between complementary artifacts. In MEET there is a very strong dependency between the smartphone application, the music player, and the situated display as none of them can work independent of the other. The tablet can however be removed without losing crucial functionality but does nothing on its own. In AirPlayer there is also a strong dependency between artifacts as no control of the music is implemented outside the smartphone application. The point is that it can be useful to consider the dependencies of complementary artifacts. Not only in the scope of the multi-artifact system but also in relation to the artifact ecologies involved. In AirPlayer all the artifacts belongs to the ecology of a single person as only one smartphone application is allowed at any time. MEET on the other hand is by design dependent on artifacts from several personal artifact ecologies.

D. Multi-user

The last concept is different from the others, as it addresses the users instead of the artifacts. Whether a system is designed for single or multi-user interaction is not surprisingly an important factor. What it means to include multiple users in terms of artifact ecologies is that the multi-artifact system spans more than one personal artifact ecology and that all involved users' ecologies intersect. MEET is for instance designed specifically for a social context with several simultaneous users. Each user has an artifact ecology, which their smartphone is a part of. When they arrive and connect their smartphone to the player the situated display and music player becomes part of each user's artifact ecology as well. Even though each smartphone at this point is part of the same multi-artifact system, they are not part of any other user's artifact ecology.

The new possibilities for designing multi-user interactions is one strength of multi-artifact systems. MEET for example, has no inherent upper limit on the number of simultaneous users by design. The possibilities do however come with a price. Just as multi-artifact systems adds an extra layer of complexity to single-artifact interaction, so does multi-user interaction. The idea of the movement feature in AirPlayer is an example where whenever there is only one person present there is no problem. Difficulties however arise if more people want to use the feature simultaneously. What happens if two persons, with different music following them, enter the same room? Rules could of course be made to cope with this problem, and as it may seem trivial to always take the number of intended users into account, it is important to do more to understand the multi-user dynamics in artifact ecologies.

E. Summary

We have analyzed the two multi-artifact music systems, MEET and AirPlayer and have identified four concepts of multi-artifact interaction: *Plasticity*, *migration*, *complementarity, and multi-user*. Fig. 4 shows an overview of where the concepts were identified in the two systems.

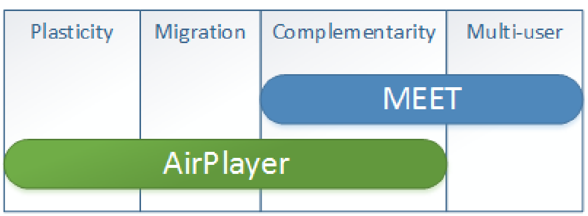


Figure 4. Utilization of discussed concepts in the two systems.

We do want to stress that the concepts should not be seen as individual solutions. There lies great opportunity in combining the concepts as was also evident in our analysis. Partial, distributing, and aggregating migration can furthermore be used to switch between complementary artifact compositions.

VII. CONCLUSION AND FUTURE WORK

The work in understanding artifact ecologies becomes important as they evolve and the relationships among artifacts become more complex. Previous work has focused on the composition and dynamics of artifact ecologies on a very high abstraction level. What we have done is to start an articulation of the sub-systems of artifact ecologies on a level in between the interaction with single artifacts and the understanding of the ecologies in their entirety. The four identified concepts of multi-artifact systems, i.e., *plasticity, migration, complementarity, and multi-user* can help obtain a more fine-grained understanding of artifact ecologies. One future step that we are already looking into is to create a clearer picture of the three layers of artifact ecologies possibly through a reference framework. An obvious next step would furthermore be to get a deeper understanding of the identified concepts with the ultimate goal of creating design guidelines for multi-artifact systems that do not only work well in isolation, but fits into an artifact ecology.

REFERENCES

[1] L. Balme, A. Demeure, N. Barralon, J. Coutaz, and G. Calvary, "CAMELEON-RT: a Software Architecture Reference Model for Distributed, Migratable, and Plastic User Interfaces," Proc. Second European Symp. Ambient Intelligence (EUSAI 2004), Springer, 2004, pp. 291-302, doi:10.1007/978-3-540-30473-9_28.

[2] S. Berti, F. Paternò, and C. Santoro, "A Taxonomy for Migratory User Interfaces," Proc. 12th Int. Workshop on Interactive Systems. Design, Specification, and Verification (DSVIS 2005), Springer, 2005, pp. 149-160, doi: 10.1007/11752707_13.

[3] S. Bødker and C. Klokmose, "The Human-Artifact Model: An Activity Theoretical Approach to Artifact Ecologies," Human-Computer Interaction, vol. 26, no. 4, 2011, pp. 315-371, doi: 10.1080/07370024.2011.626709.

[4] S. Bødker and C. Klokmose, "Dynamics in artifact ecologies," Proc. Nordic Conference on Human-Computer Interaction: Making Sense Through Design (NordiCHI '12), ACM Press, 2012, pp. 448-457, doi:10.1145/2399016.2399085.

[5] F. Capra, "The Web of Life," Anchor Books, New York, 1996.

[6] J. Forlizzi, "How Robotic Products Become Social Products: An Ethnographic Study of Cleaning in the Home," Proc. ACM/IEEE International Conference on Human-robot Interaction (HRI '07), ACM Press, 2007, pp. 129-136, doi:10.1145/1228716.1228734.

[7] L.E. Holmquist, "Ubiquitous Music," Interactions – Ambient intelligence: exploring our living environment, vol. 12, no. 4, July + August 2005, pp. 71-ff, doi: 10.1145/1070960.1071002.

[8] H. Jung, E. Stolterman, W. Ryan, T. Thompson, and M. Siegel, "Toward a Framework for Ecologies of Artifacts : How Are Digital Artifacts Interconnected within a Personal Life ?," Proc. Nordic Conference on Human-Computer Interaction: Building Bridges (NordiCHI '08), ACM Press, 2008, pp. 201-210, doi: 10.1145/1463160.1463182.

[9] E. Lenz, S. Diefenbach, M. Hassenzahl, and S. Lienhard, "Mo. Shared music, shared moment," Proc. Nordic Conference on Human-Computer Interaction: Making Sense Through Design (NordiCHI '12), ACM Press, 2012, pp. 736-741, doi: 10.1145/2399016.2399129.

[10] L. Liikkanen, C. Amos, S.J. Cunningham, J.S. Downie, and D. McDonald, "Music interaction research in HCI: let's get the band back together," CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12), ACM Press, 2012, pp. 1119-1122, doi: 10.1145/2212776.2212401.

[11] J. Rekimoto, "Multiple-computer user interfaces: "beyond the desktop" direct manipulation environments," Proc. CHI '00 Extended Abstracts on Human Factors in Computing Systems (CHI EA '00), ACM Press, 2000, pp. 6-7, doi: 10.1145/633292.633297.

[12] L. Terrenghi, A. Quigley, and A. Dix, "A taxonomy for and analysis of multi-person-display ecosystems," Personal and Ubiquitous Computing, vol. 13, no. 8, November 2009, pp. 583-598, doi: 10.1007/s00779-009-0244-5.

Following a Robot using a Haptic Interface without Visual Feedback

Ayan Ghosh
Materials and Engineering Research Institute,
Sheffield Hallam University.
Sheffield, United Kingdom
ayan.ghosh@student.shu.ac.uk

Lyuba Alboul, Jacques Penders
Materials and Engineering Research Institute,
Sheffield Hallam University.
Sheffield, United Kingdom
{l.alboul, j.penders} @shu.ac.uk

Peter Jones
The Department of Humanities,
Sheffield Hallam University.
Sheffield, United Kingdom
p.e.jones@shu.ac.uk

Heath Reed
Art and Design Research Group,
Sheffield Hallam University.
Sheffield, United Kingdom
h.reed@shu.ac.uk

***Abstract*— Search and rescue operations are often undertaken in smoke filled and noisy environments in which rescue teams must rely on haptic feedback for navigation and safe exit. In this paper, we discuss designing and evaluating a haptic interface to enable a human being to follow a robot through an environment with no-visibility. We first briefly analyse the task at hand and discuss the considerations that have led to our current interface design. The second part of the paper describes our testing procedure and the results of our first informal tests. Based on these results we discuss future improvements of our design.**

Keywords-human robot interaction; haptic interface; support for no-visibility/visually impaired

I. INTRODUCTION

In this paper, we discuss designing an interface to enable a human being to follow a robot (as shown in Figure 1). A vital pre-condition for successful human-robot cooperation in such circumstances is that the human trusts and has confidence in the robot. Trust and confidence are complex matters, which we have explored in more detail in [14]. In this paper, we focus on designing interfaces for following a robot and make a first attempt to evaluate the designs.

A. No-visibility

Being guided along an unknown path without visual feedback poses several challenges to a human being, in particular if the guide is a robot.

Contrary to popular prejudice, search and rescue operations are undertaken only when the ground is relatively easily passable [13]; the major problem however, is that the environment is smoke-filled and noisy. Rescue teams therefore must rely on haptic feedback for exploration, navigation and safe exit. However, because of the lack of visual (and auditory) feedback, humans get easily disorientated and may get lost. Robots with a range of sensors on board might be helpful for such conditions. In addition to search and rescue, there are everyday situations where vision and audition are problematic, for instance, a visually impaired person trying to navigate a busy street. Though robots are very promising, the issue of being guided by a robot is largely open and has not received much attention yet.

Figure 1. The Handle.

Young et al. [18] describe walking a robot using a dog-leash. They note that leading a robot consists of a delicate interplay between the human leader and the robot requiring ongoing communication and interaction. This includes (for both the robot and the human) monitoring the other's movement direction and speed [18]. The dog-leash is used in conditions of good visibility and a relatively low level of environmental noise. The monitoring heavily relies on visual and aural feedback i.e., the eyes and ears of the human.

However, lacking visual and aural feedback hampers orientation and causes significant stress for rescue workers as well as for the visually impaired; in addition it constitutes a significant obstacle when aiming to cater for trust and confidence. Nevertheless, psychological research has demonstrated, contrary to early assumptions and common prejudice, '*the presence of a comparable set of spatial abilities in people without vision as can be found in those with vision*' [5]. Bremner and Cowle [1][15] note: the

 ISBN: 978-1-61208-325-4

senses touch, proprioception, vision, and occasionally audition, '*convey information about the environment and body in different neural codes and reference frames*'. Research has also highlighted the extraordinary speed and sensitivity of the haptic sense [8]. This gives enough reasons to explore how to make better use of the haptic sense. Eventually, a well-designed haptic interface suitable for guidance in no-visibility conditions might also be useful in normal conditions and may free the visual sense and related mental resources so that they can be used for other tasks.

B. Navigation and following

Leading a robot is far from a simple physical locomotion problem [18]. However, making a robot lead a person raises considerable additional issues, concerning the degree of autonomy granted to the robot and the type and extent of control exerted by the human. Based on our analysis of the interaction between a visually impaired person and a guide dog we distinguish between locomotion guidance and navigation. While the visually impaired human handler determines global navigation (i.e., final destination and en-route decision points) the guide dog provides locomotion guidance between these decision points; as it can be seen in Figure 2. Locomotion guidance is effected through a simple haptic interface between dog and handler - that is a rigid handle held by handler and attached to the dog's harness.

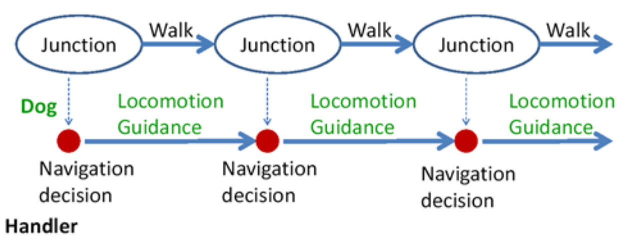


Figure 2. Handling a guide dog/robot; task analysis[14].

Inspired by this, the current paper has the focus on locomotion guidance or simply following a robot in a safe manner. However, we leave the questions on confidence and trust for future.

The paper is organised as follows: after a brief literature (Section II) review, we discuss in Section III, the design presumptions and considerations, which led to the implementation of the final interface (shown in Figure 1). In Section IV, we describe our preliminary and informal test trials. We finish with a discussion on the open issues.

II. Literature Overview

Literature on experiences of human subjects with human-robot interaction in low-visibility is rather sparse. The Guardians project [13] pioneered a group of autonomous mobile robots assisting a human rescue worker operating within close range. Trials were held with fire fighters and it became clear that the subjects by no means were prepared to give up their procedural routine and the feel of security provided: they simply ignored instructions that contradicted their routines.

There are several works on robotic assistance to the visual impaired. Tachi et al. [16] developed a guide-dog robot for the visually impaired, which leads the person. The robot tracks the follower using active sonar, and the follower wears a stereo headset, which provides coded aural feedback to notify whether the follower is straying from the path. There is no means to communicate to the robot, and the follower must learn the new aural-feedback code: the robot serves as a mobile beacon that communicates with the headset.

Allan Melvin et al., [12] developed a robot to replace a guide dog; however the paper does not extensively report trials with users. The GuideCane [17] is a cane like device running on unpowered wheels, it uses Ultra Sound to detect obstacles. The follower has to push the GuideCane - it has no powered wheels- however it has a steering mechanism that can be operated by the follower or operate autonomously. In autonomous mode, when detecting an obstacle the wheels are steering away to avoid the obstacle. The GuideCane has been tested with 10 subjects three of whom were blind and cane users, the other seven were sighted but blindfolded. Basic conclusion: '*walking with the GuideCane was very intuitive and required little conscious effort*', unfortunately nothing more is reported on the subjects' experience.

The robotic shopping trolley developed by Kulyukin [4][11] is also aimed at the visual impaired. This trolley guides the (blind) shopper - who is holding the trolley handle - along the aisles into the vicinity of the desired product. The locomotion guidance is fully robot driven but restricted to navigating the aisles; the emphasis is on instructing the shopper how to grab the product using voice instructions.

In the current paper, we restrict the use of a haptic interface (no aural or visual feedback) to locomotion guidance only. By simplifying the task, we are able to take the first step towards evaluating the subject's performance, while following the robot. The future aim is to combine the observed performance of the subjects with assessing their confidence in technology.

III. Robotic Guide

A. Design presumptions

Our final aim is to design a system and interface that allows skilled and successful guidance, enhancing human trust and confidence. We expect that a key dimension of the skillset of the human follower is the ability to 'read' the whole situation in relation to the relevant programme of action [6][7]. The aim is for *transparent technology*; technology that is so well fitted that it becomes almost invisible in use' [2]. In contrast, an 'opaque technology' is 'one that keeps tripping the user up, and remains the focus of attention even during routine problem-solving activity' [2].

 ISBN: 978-1-61208-325-4

The classic illustration of 'transparent technology' in this sense, and of particular relevance to our own study, was the use of a cane by a blind person (or 'cane traveler') for navigational purposes [2].

B. *Mechanical interface: design considerations and history*

A first step towards this aim is to build an interface that will lead the follower along a safe path. The safest path for the follower is a path that the robot already has traversed; thus the follower should follow the trail of the robot exactly. Hence our experiments, reported below, look at the following behaviour of the follower in terms of the ability to closely match the live path of the robot.

Obviously, in order to be able to follow the robot, the follower needs to know where the robot is relative to his/her current position and orientation. Initially our project looked at three distinct interfaces: a wirelessly connecting device for instance a Nintendo Wii, a short rope/rein or leash and a stiff handle. A major problem for any wireless device lies in how to indicate the position of the robot with respect to the follower. A rope does indicate the direction of the robot but only when there is no slack. Young et al. [18] use a spring-loaded retractable leash design (popular with dogs), which keeps the leash taut; the retracting mechanism however obscures the length of the leash and thus the distance between the robot and the follower is not known. Our final choice has been for a stiff handle via which the position (direction and distance) of the robot is immediately clear to the follower.

C. *Interaction with a Stiff interface:*

We tried a stick held in one hand mounted on a disc with unpowered omni-directional wheels (as presented in Figure 3). Basically, the disc would be set into motion by the person holding the stick. The omni-directional wheels made the disc easy manoeuvrable in any direction (on the floor). However, when holding the stick blind folded, a lack of accuracy in sensing the direction has been noticed; several subjects immediately put their second hand on the stick to compensate. Our observation of a lack of accuracy of a one handed hold is in line with experiences in using a white cane. Visually impaired people using a white cane do hold the cane in one hand but they also apply a special grip (for instance stretched the index finger) and/or keep the elbow touching the body. We note that manipulating our disc is not as easily as handling a white cane.

From this we concluded that a crutch like design of the handle, in which the stick is fixed on the lower arm, is preferred.

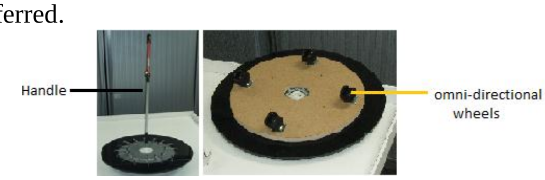


Figure 3. Hand held stick on a disc with omni-directional wheels.

D. *Implementing the handle (stiff Rein) on the robot*

Based on these conclusions, a simple crutch-like prototype with a ball-free mechanism at the base (as presented in Figure 4) was developed to enable some initial experimentation. The pilot studies have revealed that, there have been instances such as:

Figure 4. Ball-free mechanism at the Base.

- The follower did not feel safe following the robot (as presented in Figure 5). Obviously we can judge the path of the follower as safe when it closely matches the path of the robot.
- The follower lost track of the orientation (heading) of the robot, though its position was clear. As a consequence, the follower did not feel comfortable following the robot at the turns. The handle delivered an abrupt tug to the follower at the point of the turns.

Figure 5. Unsafe path, the follower gets deviated too much off the course.

These findings led to the design of a third prototype to ensure safety, comfort and rigidity. The prototype consists of a mechanical feedback spring system at the base, as presented in Figure 6. The spring system allows rotation of the handle on the horizontal plane. When the spring system has zero tension, the handle is aligned with the center line of the robot. When the handle is being rotated, the spring system induces tension on the handle, which increases with the rotation angle. The system also comes with a pin enabling to nullify the action of the springs, giving us the option to carry out a comparative study between a flexible joint and a fixed joint. Thus, this handle provides two testing options:

- *The handle is attached in a fixed joint (rigid)*: meaning the handle is fixed at base using the pin.

- *The handle is attached with a flexible joint (spring)*: meaning the handle can rotate in the horizontal plane, and rotation induces tension on the handle.

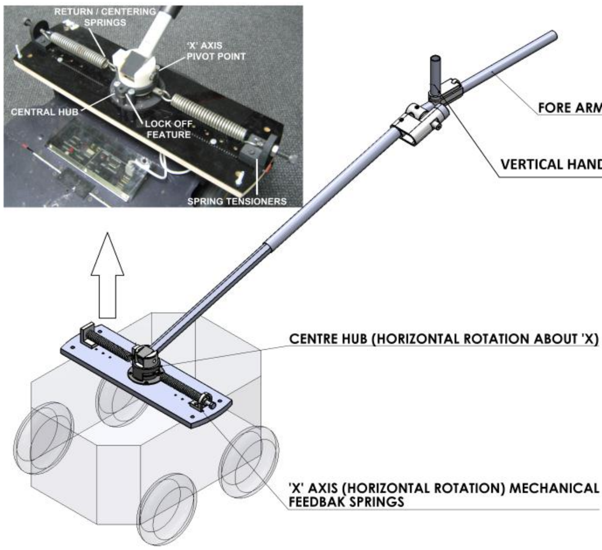


Figure 6. Handle with spring system.

E. *Robot and sensors*

The handle has been mounted on a Pioneer-3AT 4-wheel robot. In the experiments reported below, the robot was autonomously navigating fixed trajectories while being supervised by an operator. The operator was able to stop/start the robot remotely using a developed Java application [9]. The robot operated with a linear speed of 0.6m/s and the angular speed was set at 0.5 rad/s (at the turns).

At all time, the walking pattern of the follower was being observed and the degree of displacement of the follower with respect to the center line of the robot was being recorded using a Hokuyo Laser Range Finder, which was fixed exactly at the middle of robot's rear bumper. Data collection proceeded at a speed of 10Hz or 10 observations per second. The positions of the robot at every instance of time were measured by odometry sensors. The data was sent to the operator's workstation using a Lantronix 802.11g WiPort modem.

IV. ROBOTIC EVALUATION GUIDELINES

In designing and interpreting our preliminary experimental studies, we were guided by the theoretical perspective of developing the robot guide as 'transparent technology' [2] [3]. And, the primary evaluation purpose was to test usability: whether a person could easily follow the robot.

A. *Testing Protocol*

We studied the effect of two different settings of the stiff interface on the following behaviour of right-handed participants. On each of the trials, the subjects were asked to use the stiff handle in one of the following modes:

- *The handle attached in a fixed joint (rigid)*
- *The handle attached with a flexible joint (spring)*

The overall aim of the study is to evaluate the use of an autonomous robot guide. However, autonomous behaviour can occur in many variants; for our study, we confined the robot to five pre-programmed repeatable behaviours. Thus, the robot was made to move autonomously in one of the following pre-programmed trajectories below:

- *path A*: Straight line (approximately 8 meters).
- *path B*: Straight line (approximately 5 meters) + sharp turn (right/left) + straight line (approximately 3 meters).
- *path C*: Straight line (approximately 5 meters) + gentle turn (right/left) + straight line (approximately 3 meters).

When the robot moves in a straight line, the set linear speed is inspired by the normal walking speed of a person. However, for setting the robot's angular speed we do not have an intuition; therefore we designed a smooth turn (close to 45 degrees) and a sharp turn (close to 90 degrees).

Our preliminary and informal tests were carried out with team members (four) as subjects; each of them performing 8 trials for each of the paths A, B and C, with different handle settings. Subjects were blindfolded and asked to put headphones on. Before the commencement of each trial, the handle was attached to the subject's forearm and a gentle pat was the pre-arranged haptic signal from the experimenter, used to indicate the start of each trial. For each trial we monitored the following:

- the position coordinates (odometry sensors) of the robot in the experimental space, at a frequency of 10 Hz .
- the degree of displacement of the subject from the trajectory of the robot.

The data collected were used to examine the spatial correspondence of the robot's path and the follower's path.

B. *Experimental results*

Robot following path A:

Our first trial with each subject aimed to observe how the person follows the robot. The handle is mounted in the middle of the robot, while the crutch like part of the handle is attached to the right fore-arm of the follower (right-handed) thereby making him/her stand about 15-20 cm left of the center line of the robot (as presented in Figure 7). In the figures below, we show reconstructions of the paths of the robot and the follower across several trials. The reconstruction is based on the data collected (10 Hz) on board of the robot. The movements (straight/left/right) of the robot and follower are shown in the diagrams.

The robot is around a meter (length of the handle) in front of the follower. So while the robot starts at time t_0 at position (0,0) the follower is at time t_0 at position (-1,0). Figure 8 shows a reconstruction of the straight path (path

A). When the path is straight, there is no impact of handle settings (fixed or sprung joint) on the following behaviour: the follower follows the robot, slightly (15-20 cm) off the robot's centre.

Figure 7. Position of the follower at the start.

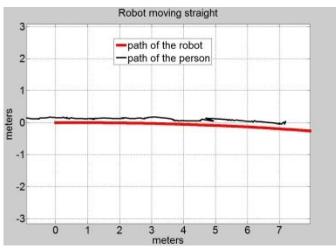


Figure 8. Reconstruction of the paths of robot and follower moving in a straight-line.

Robot following path B:

Figure 9 and Figure 10 show a reconstruction of the paths while the robot takes a sharp turn to the left. It is visible across the trials that there is a very obvious difference between the follower's experience with fixed joint (Figure 9) and the sprung joint (Figure 10) and the impact of these two different handle settings on the follower's following behaviour. When the joint is fixed as in Figure 9 the follower is forced across the centre line of the robot. The follower gets deviated about 0.5 m of the life path of the robot. With the flexible joint this effect is rather minimal and there is a higher degree of matching of paths (as Figure 10 shows).

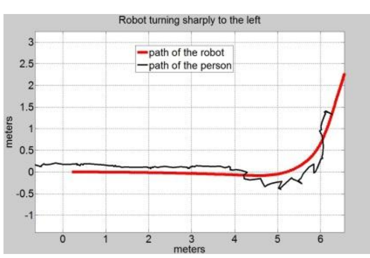


Figure 9. Reconstruction of the paths of robot and follower with sharp turn (fixed-joint).

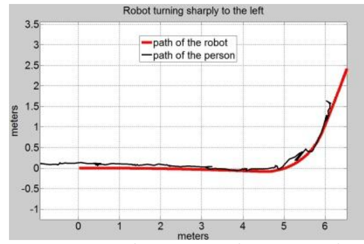


Figure 10. Reconstruction of the paths of robot and follower with sharp turn (sprung-joint).

Robot following path C:

Figure 11 shows the reconstruction of the paths of both robot and follower, while the robot takes a gentle turn. In this case, when the robot turns, there is also a clear, but smoother deviation of the follower's path from the path of the robot.

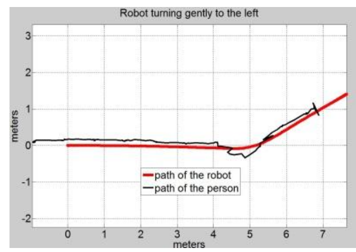


Figure 11. Reconstruction of the paths of robot and follower with gentle turn (fixed-joint).

Turning right versus turning left

It became evident from our experiments that there is an acute difference in the following behaviour when the robot is turning right and when the robot is turning left. On right turns, the follower's path deviates considerably more from path of the robot (at the point of turn) than on left turns, compare Figure 9 with Figure 13.

The follower is holding the handle in the right hand; when the robot is taking a right turn, the crutch like handle pushes the follower's arm towards the body of the follower. This is forcing the follower to step out; at the same time the initial 'inertia' of the follower causes slippage of the robot meaning that Figure 13 and Figure 14 also include a slippage error.

Figure 12. The body posture of a person during left (left) and right (right) turn.

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-325-4

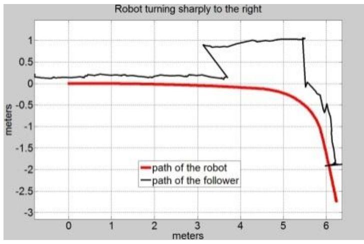


Figure 13. Reconstruction of the paths of robot and follower with sharp right turn (sprung-joint).

Contrarily, during a left turn the arm has much more freedom for movement and the following behaviour looks more comfortable. Figure 12 shows the body postures of a person when the robot starts turning right and left. These effects are persistent during gentle turns as well (as shown in Figure 14).

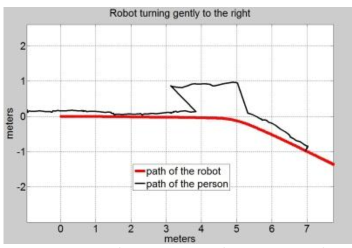


Figure 14. Reconstruction of the paths of robot and follower with gentle right turn (sprung-joint).

V. DISCUSSION

The findings of the experimental trials raise a number of issues about the design of the handle and user experience that deserve further investigation. First of all, it seems clear that when the handle is attached with a flexible joint (spring) the follower's path better matches the path of the robot; there is only little displacement of the human follower from the robot's trail. For right turns, deviations start very abrupt, but remain smaller with the sprung-joint. In the turns the follower is exerting some force on the robot and this causes the robot to slip and maybe slide. The reconstructed paths in figures 9-14 are based on odometry data and will contain some error, nevertheless the overall patterns can be recognised in the videos taken.

The flexible handle setting allows for a build-up of tension within the spring mechanism in real time, meaning that the forces on the subject accumulate gradually, thereby causing a delay between the start of the robot's turn and the follower reacting to it (the start of the subject's turn). That delay makes for a smoother turn and one that is more accurate spatially, however, it leaves open how immediately and accurately the follower is alerted of the movements of the robot through the haptic interface.

In terms of the subjective experience of the follower, our initial anecdotal evidence suggests that the flexible handle setting affords a smoother and more comfortable guided experience, although the firmer and more abrupt tug delivered by the inflexible handle may give the handler a keener awareness of spatial orientation and location.

Future experiments will have more formal layout and will include questionnaires in order to capture the subjective experiences. Also, we will have to compare right and left handed subjects in order to confirm our intuition that on a left turn a left handed person is also forced to step out and mirrors the pattern of a right turn by a right handed person.

Future work will concentrate on refining the objective and subjective measures of path correspondence and examine to what extent following can be seen as a learnable skill, with the handle becoming 'transparent technology' and helping in 'human-technology symbiosis' [10].

VI. CONCLUSION

In this paper, we have presented a haptic interface attached to an autonomous robot for locomotion guidance. We have reported on a small scale experimental study of different settings of the interface. We have learned that the feedback spring mechanism at the base of the interface created a quite different feel to the task of following the robot without any visual and audio feedback, giving more safety and comfort.

ACKNOWLEDGMENT

This research was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant no. EP/I028765/1.

The authors also would like to thank Alireza Janani for providing measurements for human and robot paths.

REFERENCES

[1] A. J. Bremner, and D. Cowie, "Developmental Origins of the Hand in the Mind," and "The Role of the Hand in the Development of the Mind" (in [15], 2013).

[2] A. Clark, "Natural Born Cyborgs:Mind,Technologies and The future of Human Intelligence," Oxford: Oxford University Press, 2003.

[3] A Clark, "Supersizing the Mind: Embodiment, Action, and Cognitive Extension," Oxford: Oxford University Press, 2011.

[4] C. P. Gharpure, and V. A. Kulyukin, "Robot-assisted shopping for the blind: issues in spatial cognition and product selection," Intelligent Service Robotics, 2008, Volume 1, Number 3, 237-251, DOI: 10.1007/s11370-008-0020-9.

[5] R.G. Golledge, R. L. Klatzky, and J. M. Loomis, "Cognitive mapping and wayfinding by adults without vision," in J Portugali (ed.) The Construction of Cognitive Maps, Kluwer Academic Publishers: 215-246, 1996.

[6] R Harris, "Signs, Language and Communication," London: Routledge, 1996.

[7] R Harris, "The integrational conception of the sign", in Integrationist Notes and Papers 2006-2008, Sandy:Bright Pen, 61-81, 2009.

[8] M. A. Heller, W. Schiff, and L. Erlbaum, "The Psychology of Touch," Inc. 1991.

[9] A. Janani, A. Holloway, H. Reed, and J. Penders, "Design of a Mechanical Impedance Filter for Remote Haptic Feedback in Low-Visibility Environment," TAROS, 2013.

[10] P. Jones, A. Ghosh, J. Penders, and H. Read, "Towards human technology symbiosis in the haptic mode,". In: International Conference on Communication, Media, Technology and Design, Famagusta, North Cyprus, 2-4 May 2013. 307-312, 2013.
[11] V. Kulyukin and A. Kutiyanawala, "Accessible shopping systems for blind and visually impaired individuals: Design requirements and the state of the art," The Open Rehabilitation Journal, 2010.
[12] A. Melvin, A. Prabu, B. Nagarajan, R. Bukhari, and Illia, "ROVI: a robot for visually impaired for collision- free navigation," Proceedings of the International Conference on Man-Machine Systems (ICoMMS 2009).
[13] J. Penders, L. Alboul, U. Witkowski, A. Naghsh, J. Saez-Pons, S. Herrechtsmeier, and M. El-Habbal, "A robot swarm assisting a human firefighter," Advanced Robotics, 25:93-117, 2011.
[14] J. Penders, P. Jones, A. Ranasinghe, and T. Nanayakara, "Enhancing trust and confidence in human robot interaction," In: UKRE, Sheffield, 25-3-2013.
[15] Z. Radman, "The Hand, an Organ of the Mind," MIT Press, 2013.
[16] S. Tachi, R. W. Mann, and D. Rowell, “Quantitative comparison of alternative sensory displays for mobility aids for the blind,” IEEE Trans. Biomedical Engineering, vol. 30, no. 9, pp. 571–577, Sep. 1983.
[17] I. Ulrich, and J. Borenstein, "The GuideCane-applying mobile robot technologies to assist the visually impaired," Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on Issue Date: Mar 2001 Volume: 31, Issue:2 On page(s): 131 - 136 ISSN: 1083-4427.
[18] J. E.Young, Y. Kamiyama, J. Reichenbach, T. Igarashi, and E. Sharlin, "How to Walk a Robot: A Dog-Leash Human-Robot Interface," In: International Symposium on Robot and Human Interactive Communication, RO-MAN, pp. 376–382 (2011).

Posture-Angle Perception and Reproduction Characteristics with Wrist Flexion/Extension Motions

Yoshihiko Nomura, Tatsuya Ito

Graduate School of Engineering
Mie University
Tsu, Japan
nomura @ mach.mie-u.ac.jp, ito @ mach.mie-u.ac.jp

***Abstract*—Comparing active and passive modes with wrist flexion/extension motions, the authors examined the posture angle perception and reproduction characteristics through psychophysical experiments using a mechanically haptic wrist interface: mean values and standard deviations of the perceptual/reproduced angle errors were obtained, and were examined by ANalysis Of VAriance (ANOVA). The characteristics can be applied to wearable haptic interfaces employing kinesthetic sensations in a form of an instruction scheme referred to as a "restrained instruction": the restrained instruction is based on the idea that the just-noticeable small amount of externally applied forces being would be enough for learners to trigger voluntary motions with their body elements.**

Keywords-motion instruction; wrist; flexion; extension; perception: reproduction

I. Introduction

In recent years, exoskeleton robotic-suits and rehabilitation-systems have been developed in various forms such as full-bodies [1][2], gait rehabilitations [3], arms [4][5][6]. It was also realized as grounded instruction systems [7]. They exerted large enough forces for person body elements to be moved passively

Contrary to these power assisting systems, pose and kinesthetic senses embodied in human bodies can be utilized for motion instruction in the form of body-worn haptic interfaces. As with the kinesthetic-sensation-based motion instruction schemes, J. Iqbal et al. presented a prototype of hand exoskeleton-type finger motion-assisting device for accomplishing common daily life activities, and showed some optimization algorithms for mechanical design [8][9]. Muscle spindles are so sensitive to notice muscle contractions, and, therefore, noticeable threshold levels of human joint flexion/extension movements were reported to be a very small level of less than 0.1 degrees [10][11]. Thus, the thresholds were very small, which give us a suggestion that we need not a large amount of external stimuli for notifying us of the movements.

Considering this point, the authors have been developing interfaces that are characterized by featuring a novel instruction scheme, i.e., just-noticeably small-force scheme: it is assumed that the just-noticeable small external forces would be enough to trigger learners to voluntarily move their muscles. We call the voluntary motion as "active", and the scheme as a "restrained instruction". The restrained instruction scheme has an advantage of small power, and makes systems compact and light weighted. Furthermore, the actively-inspired-motion-based instruction scheme is expected to be effective for learning motions compared to the passively-forced-motion-based instruction scheme. Here, as for the with/without human-initiative effects on the performances, some results were summarized by Proske and Gandevia [12]. As for the effects of muscle conditioning on position sense at the human elbow flexors, T. J. Allen et al. reported in the left/right forearms matching task that there were not significant differences with respect to the effect of the sense of effort [13]. Contrary to this, S. C. Gandevia et al. showed a role for efferent outflow signals in the mean that motor commands contribute to human position sense [14], and the effects of human initiative were not confirmative. Thus, paying attention to the initiative factor, the authors have studied on absolute-angle perceptual characteristics with the wrist flexion/extension using a mechanically haptic wrist interface as well as the elbow joint [15].

Learning some specific postures is one of the important processes in motion learning processes: for examples, still-posed postures are essential for some exercises such as "yoga" positions, and, even in dynamic motions, instantaneous postures at the motion-phase transitions are also crucial. Especially in the case of the restrained instruction, we shall feel it difficult to notice the external exhilarated motions than in the other case of forcible instructions. Even in such cases, we would be able to notice the joint angle changes between the before-and-after instructions if the changes were larger than Just Noticeable Differences (JNDs). Therefore, the JNDs are useful for learning some specific postures, and were examined as well as the mean errors in this paper.

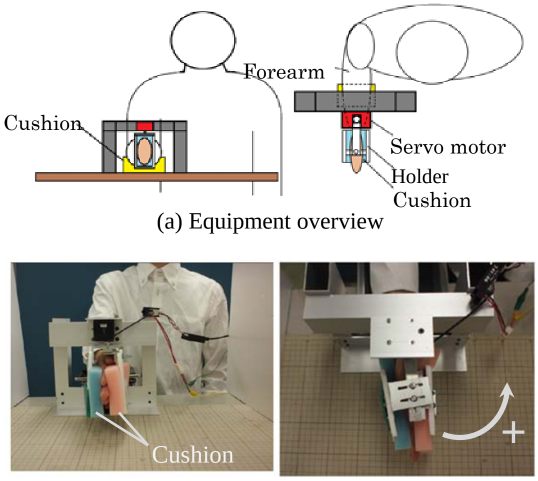
Section II explains the psychophysical experimental method, i.e., apparatus and procedures, for examining the wrist flexion/extension angle perceptual characteristics. Next, Section III first presents the experimental results, and examines the systematic and random errors: the former is given by the mean error and the latter is by the standard deviations. Both the errors suggest reproducibility in the still-posed posture reproductions after motion learning procedures. Finally, Section IV addresses the conclusion and future work.

II. Experimental Method

This section describes the psychophysical experimental method, i.e., apparatus and procedures, for examining the wrist flexion/extension angle perception/reproduction performances in the extension/flexion motions. Here, assuming that there might be a difference in the performances with respect to subject initiative, the authors introduced two levels of a subject initiative factor to make a comparative study: one level is an active haptic scheme, and the other one is a passive haptic scheme.

A. Apparatus

A simple wrist-bending apparatus was designed to carry out a psychophysical experiment for examining the wrist flexion/extension angle perceptual characteristics. Figure 1 shows the experimental apparatus. A servo-motor was attached to a lower side of the gate-shaped aluminum frame. The servo-motor exerts torques, and makes subject's wrist to forcibly do flexion/extension motions via cushion-buffered-holders.



(a) Equipment overview

(b) Front view (c) Top view (Flexion angle, 0°)

Figure 1. Experimental system setup

B. Procedures

1) Conditions: In this experiment, a back-and-forth motion from an initially instructed before-bend angle, via another angle that were given as mask stimulus, to the after-bend angle was tested. Relating to the back-and-forth motion, kinesthetic sense characteristics, i.e., wrist-bend perceptual characteristics, were examined.

Here, focusing attention to the motion initiative factor, the authors introduced two levels: one is a passive haptic mode, and the other is an active haptic mode. The passive haptic mode (Pa) is considered to be a representative passive-based instruction. The subject does not take initiative, but the actuator forcibly rotated the subject's wrist. An ability to notice the deviations of the after-bend angles from the before-bend angles was examined: perceptual errors of the differences between the initially instructed angles and the returned angles were examined in this mode. The active haptic mode (Ac) represents a realization concept of "the restrained instruction scheme" proposed in this paper. The subject takes initiative, and voluntarily activates the wrist flexion/extension. An ability to reproduce the initially instructed angles after an active bending process was examined in this mode. That is, the reproduction errors between the initially instructed before-bend angles and the returned after-bend angles, to which subjects voluntarily bent their wrist so as to reproduce the initially instructed angles, were examined.

In addition to the initiative, the subject and the interaction factors, the other two factors, i.e., an initial angle factor, and an angular velocity factor were examined. The levels of the initial angle factor were (1) the straight wrist condition, i.e., 0°, (2) a medium level of dorsiflexion condition of -30°. The other levels of the angular velocity factor were (1) a relatively higher speed of 30°/s, (2) a lower speed of 10°/s.

Right handed four male subjects aged 23 to 59 participated in the experiment. Thus, the total number of factor level combinations was given by 32(=2×2×2×4). As described in the next section, for each of the factor level combinations, tasks were repeated by 10 runs, and, therefore, the total number of runs was 320(=32×10).

2) Task and Procedure: The procedure was composed of pre-steps and perceptual steps as in the followings.

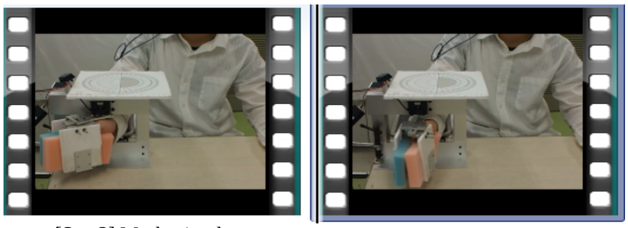
[Pre-step 1] The subject sat on a chair, and his right hand was held tight to the equipment via the opposing cushions, so that his forearm being horizontal and their upper arm being vertical.

[Pre-step 2] The subject closed his eyes, and a white noise sound was applied to the subject via headphones for masking any sound cues on the perception.

The perceptual steps were implemented for both the passive and active modes:

a) Passive mode: This mode was carried out by the following procedures (see Figure 2).

Preparatory step [Step1] Test target stimulus acceptance.



[Step3] Mask stimulus acceptance
[Step4] Pseudo-test stimulus acceptance

Figure 2. Experimental procedures in the passive mode

[Step1] Passive test target stimulus acceptance: the system forcibly bent a subject wrist from an arbitrary position to an initial angle using the servomotor. This step simulated a process of forcible initial position learning process.
[Step2] The subject was informed that his wrist angle had come to the initial position.
[Step3] Passive mask stimulus acceptance: the system forcibly bent the subject wrist to approximately -60° in the dorsiflexion direction. This process simulated a consecutively delivered passive instruction.
[Step4] Passive pseudo-test stimulus acceptance: the system, finally, forcibly returned the subject wrist to a destination angle chosen from a set of angles a little bit deviated from the initial angle: the deviations from the initial angle were -20°, -16°, -12°, -8°, -4°, 0°, 4°, 8°, 12°, 16°, 20°. Ten kinds of deviations were chosen for them and were presented. This process also simulated the second-consecutively delivered passive instruction.
[Step5] Looking at an answer board being set horizontally, the subject opened his eyes, and answered the amounts of his perceived value with the deviation of the destination angle from the initial angle: the answer board was a protractor-like scale, the fineness of which was 1° (see Figure 3). The ability to notice deviations was considered to be related to an ability for passive posture reproductions.

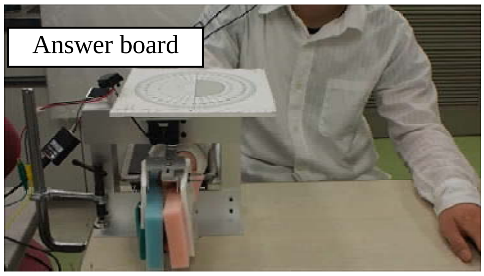


Figure 3. Answer board: subjects chose a character showing the amounts of their perceived angular difference between the before-and-after-bend angle.

b) Active mode: This mode was carried out by the following procedures (see Figure 4).

Preparatory step [Step1] Test target stimulus acceptance.

[Step3] Mask stimulus self-production
[Step4] Test stimulus self-reproduction

Figure 4. Experimental procedures in the active mode: there is no difference in the appearance between this figure and Figure 3. The crucial difference is in whether there is a subject's intention or not.

[Step1] Passive test target stimulus acceptance: the system forcibly bent the subject wrist to an initial angle using the servomotor. This step simulated a process of the forcible learning process.
[Step2] The subject was informed that his wrist had come to the initial position. Then, the electrical current of the driving servomotor was cut off, which made the subject wrist to move freely with small torque being enough to cancel frictions induced at reduction gears.
[Step3] Active mask stimulus self-production: the subject bent one's wrist up to the mechanical limit. This process simulated a process where learners themselves voluntarily change their postures.
[Step4] Active test stimulus reproduction: the subject bent one's wrist up towards the initial position. This process was another voluntarily changing process, and simulated a process to reproduce the specific postures instructed before.
[Step5] The system measured the subject wrist angle.
A series from Step 1 to Step 5 was repeated ten times.

III. Experimental Results

In this section, the experimental results are first presented, and, then, they are examined from the viewpoint of the mean errors and the standard deviations. The mean errors represent biases, and are, so-called, the systematic errors. The standard deviations represent the widths of scattering, and are, so-called, the random errors. Both the errors suggest reproducibility in the still-posed posture reproductions after motion learning procedures. The latter standard deviations will be also related to JNDs.

A. Systematic Errors

In the passive mode, the errors were defined by the perceived angle errors, i.e., the angular differences of the perceived angles from the true angles that had been deviated by -20° to 20° from the initial angles. On the other hand, those in the active mode were defined by the angular differences of the reproduced angles from the initial angles.

As for the perceived/reproduced angle errors, the factors to be tested were the subject activity, the initial angles, the angular velocities, the subjects, and the interaction. Class means represent the systematic errors, and the experimental results with the global mean and the class means for three factors are shown in Figure 5.

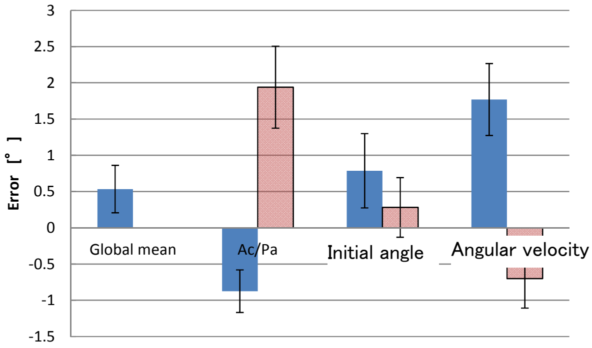


Figure 5. Experimental results of systematic errors of the reconstructed/ perceived angle errors: the main factor effects of the three factors and the global mean.(Error bars: standard errors.)

The differences from the global mean are, so-called, the main factor effects. The significances among the between-class variations of the factor effects were tested by repeated-ANalysis Of VAriances (ANOVA), as shown in Table I.

TABLE I. ANOVA Table with Perceptual Angle Errors

Factor	Item / Level	Mean [°]	Factor effect [°]	Sample size	Variation [°^{2}]	DOF	Mean square [°^{2}]	Test statis. F-val.	0.1% point (Crit. val.)	Decision
	Glob.mean	0.534		320						
Initiative	Ac	-0.876	-1.41	160	318					
	Pa	1.94	1.41	160	318					
	Sum				636	1	636	29.2	11.0	***
Initial angle	0	0.786	0.252	160	10.2					
	-30	0.281	-0.252	160	10.2					
	Sum				20.4	1	20.4	0.94	11.0	NS
Subject				320	671	3	224	10.3	5.6	***
Angular velocity	10	1.77	1.24	160	244					
	30	-0.701	1.23	160	244					
	Sum				488	1	488	22.4	11.0	***
Interaction					2450	3	815	37.4	11.0	***
Error					6750	310	21.8			
Total				320	11000	319				

NS: not significant * P<0.05, ** P<0.01, *** P<0.001

TABLE II. ANOVA Table with JND of Perceptual Angle Errors

Factor	Item / Level	Mean [°]	Factor effect [°]	Sample size	Variation [°^{2}]	DOF	Mean square [°^{2}]	Test statis. F-val.	0.1% point (Crit. val.)	Decision
	Glob.mean	2.95		320						
Initiative	Ac	1.82	-1.13	160	178					
	Pa	4.07	1.13	160	178					
	Sum				255	1	40.5	54.6	13.9	***
Initial angle	0	3.14	0.192	160	176					
	-30	2.76	-0.192	160	109					
	Sum				109	1	1.18	1.59	13.9	NS
Subject					217	3	1.00	1.35	7.5	NS
Angular velocity	10	2.97	0.025	160	107					
	30	2.92	-0.026	160	107					
	Sum				214	1	0.02	0.03	13.9	NS
Error					3.44×10^3	25	0.74			
Total				320	4.30×10^3	31				

The initiative factor (Ac/Pa) shows a significant difference with a 0.1% level. In the case of the active mode, there can be seen a tendency of systematic errors in the dorsiflexion direction, while, in the other case of the passive mode, there was an opposite tendency in the palmar flexion direction. From the viewpoint of the magnitudes, the active mode was superior to the passive mode.

The angular velocity factor also shows a significant difference with a 0.1% level of significance. The factor effects of either the initiative factor or the angular velocity factor were 1 to 2 °. The subject factor and the interaction factor also show significant differences. The remaining initial angle factor alone shows no significant difference.

The factor effects can be applied in the practical motion instruction: they enable us to estimate systematic errors and to cancel the errors.

B. Random Errors

The Standard Deviations (SDs) of the perceived/reproduced angle error variations represent the random errors, and, furthermore, can be converted an important measure of JND. The JNDs suggest some limits of accuracies of angular instructions.

Here, note that the JND represents the difference between a pair of stimuli, and is defined by the 75 % point in the psychometric curve. Therefore, if the psychometric curve is to be approximated by the normal distribution, the cumulative probability of 75 % corresponds to 0.674×SD. Thus, JNDs are approximately converted from the SDs by

$$JND = 0.674 \times SD_{sum}. \quad (1)$$

where SD_{sum} is given by the sum of the global mean SD_{gm} and the three factor effects, $E_{Ac/Pa}$, $E_{Initial\ angle}$, $E_{Angular\ velocity}$. That is,

$$SD_{sum} = SD_{\gamma m} + E_{Ac/Pa} + E_{Initial\ angle} + E_{Angular\ elocity}. \quad (2)$$

As for the converted JNDs, some of the experimental results with the global mean and the class means for three factors are shown in Figure 6.

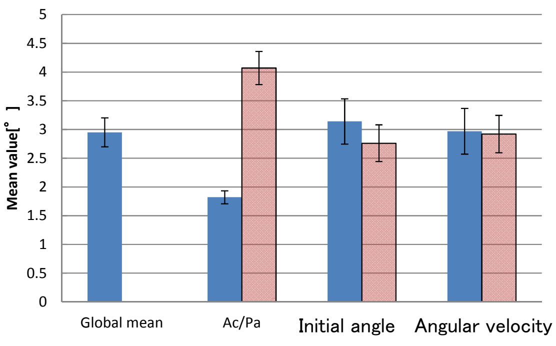


Figure 6. Experimental results of JNDs obtained from the random errors of the reconstructed/perceived angle errors: the main factor effects of the three factors and the global mean .(Error bars: standard errors.)

The significances among the between-class variations of class means were, also, tested by repeated-ANOVA as shown in Table II. We can find the initiative factor (Ac/Pa) shows significant difference with the significant level of 0.1%. Besides the initiative factor, we couldn't find any significant difference for the other factors of the initial angle, the angular velocity, and the subject.

IV. CONCLUSION AND FUTURE WORK

For establishing a new motion instructing scheme, i.e., the restrained instruction scheme, the authors have studied kinesthetic sensation with wrist bending motions.

- As for the subject's initiative, the active mode was superior to the other passive mode from both the viewpoints of the systematic errors and the random ones in an initial angle reproductions task.
- The former knowledge of the systematic errors enabled us to estimate and to cancel the errors.
- The latter results were converted into the JNDs, and the JNDs suggested some limits of accuracies of angular instructions.

In the future, we will make a profound study on the kinesthetic sensations of wrist bending motions and will extend to other joint motions.

ACKNOWLEDGMENTS

This work was supported by KAKENHI (Grant-in-Aid for Challenging Exploratory Research 25560112 from Japan Society for the Promotion of Science (JSPS)

REFERENCES

[1] T. Hayashi, H. Kawamoto, and Y. Sankai, "Control method of robot suit HAL working as operator's muscle using biological and dynamical information," IEEE/RSJ Intern. Conf. on Intelligent Robots and Systems, 2005 (IROS 2005), Aug. 2005, pp. 3063-3068.

[2] J. F. Veneman, R. Kruidhof, E. E. G. Hekman, R. Ekkelenkamp, E. H. F. van Asseldonk, and H. van der Kooij, "Design and evaluation of the LOPES exoskeleton robot for interactive gait rehabilitation," IEEE Trans. on Neural Systems and Rehabilitation Engineering, vol. 15., no. 3, Sept. 2007, pp. 379-386.

[3] C. R. Carignan, M. P. Naylor, and S. N. Roderick, "Controlling shoulder impedance in a rehabilitation arm exoskeleton," 2008 IEEE International Conference on Robotics and Automation, pp. 2453-2458, May 2008

[4] M. Mihelj, T. Nef and R. Riener, "ARMin II – 7 DoF rehabilitation robot: mechanics and kinematics," 2007 IEEE International Conference on Robotics and Automation, pp. 4120-4125, April 2007.

[5] A. Den`eve, S. Moughamir, L. Afilal, and J. Zaytoon, "Control system design of a 3-DOF upper limbs rehabilitation robot," Computer Methods and Programs in Biomedicine, vol. 89, pp. 202–214, Feb. 2008

[6] K. Kiguchi, K. Iwami, M. Yasuda, K. Watanabe, and T. Fukuda, "An exoskeletal robot forhuman shoulder joint motion assist," IEEE/ASME Trans. on Mechatronics, vol. 8, no. 1, pp.125-135, March 2003

[7] A. Gupta and M. K. O'malley, "Design of a haptic arm exoskeleton for training and rehabilitation," IEEE/ASME

Trans. on Mechatronics, vol.11, no. 3, June 2006, pp. 280-289.

[8] J. Iqbal, N. G. Tsagarakis and D. G. Caldwell, “Design of a wearable direct-driven optimized hand exoskeleton device,” 4th International Conference on Advances in Computer-Human Interactions (ACHI), Feb. 2011, pp. 142-146.

[9] J. Iqbal, N. G. Tsagarakis, A. E. Fiorilla, and D. G. Caldwell, “Design requirements of a hand exoskeleton robotic device,” 14th IASTED International Conference on Robotics and Applications (RA), Massachusetts US, Vol.664, No.81, Nov. 2009, pp. 44-51.

[10] D. L. Sturnieks, J. R. Wright, and R. C. Fitzpatrick, “Detection of simultaneous movement at two human arm joints,” Jour. of physiology, vol. 585, no.3, Dec. 2007, pp. 833-842.

[11] D. I. McCloskey, “Detection and execution of movements,” Psychol Res 55, June 1993, pp. 139-14.

[12] U. Proske and S. C. Gandevia, “The kinaesthetic senses,” J Physiology, vol. 587, no.17, Sept. 2009, pp. 4139–4146.

[13] T. J. Allen, G. E. Ansems, and U. Proske, “Effects of muscle conditioning on position sense at the human forearm during loading or fatigue of elbow flexors and the role of the sense of effort,” J Physiology, Vol.580, April 2007, pp. 423–434.

[14] S. C. Gandevia, J. Smith, M. Crawford, U. Proske and J. L. Taylor, “Motor commands contribute to human position sense,” J Physiology, vol. 571, March 2006, pp. 703–710.

[15] T. Nishimura, Y. Nomura, and R. Sakamoto, “A Restrained Torque-based Motion Instructor: Forearm Flexion/Extension-driving Exoskeleton,” Proc. Part of SPIE-IS&T Electronic Imaging, Conference on Intelligent Robots and Computer Vision XXX: Algorithms and Techniques, no. 8662-20, Feb. 2013, pp.1-8.

Position and Force-direction Detection for Multi-finger Electrostatic Haptic System Using a Vision-based Touch Panel

Taku Nakamura and Akio Yamamoto
Dept. of Precision Engineering
The University of Tokyo, Japan
Email: {taku_nkat, akio}@aml.t.u-tokyo.ac.jp

***Abstract*—This article proposes a method of detecting position and force-direction for multi-finger electrostatic haptic system using a vision-based touch panel. The system consists of a vision-based touch panel, a transparent ITO (Indium Tin Oxide) electrode on the panel, and multiple stimulators with markers. Analyzing positions of the markers with the vision-based touch panel, the proposed method can detect position and force-direction of each stimulator. Detection of applied force direction is imperative to reduce undesired stickiness of virtual walls in passive haptic rendering systems. The developed system could successfully reduced the stickiness, but its performance was limited due to the limited tracking performance of the touch panel.**

Keywords-Surface haptics; Visuo-haptic; Passive haptic system.

I. Introduction

Recently, multi-finger interaction on flat panel displays has been a hot topic in computer-human interaction. Many mobile devices and personal computers are now equipped with touch panels that allow user operations on their screens using multiple fingers. However, these devices do not fully exploit their multi-touch feature for information output, since they do not have haptic feedback. There are some systems that can provide haptic feedback on a flat panel display, such as electro-tactile display [1]–[3], string-based stimulator [4] and electromagnetic actuation [5], which however can provide stimuli to a single finger alone; rendering independent stimulus to multiple fingers have not been realized.

To realize haptic feedback to multiple fingers on a flat panel display, we have proposed on-screen multi-finger haptic feedback systems in [6] and [7], based on indirect electrostatic stimulation. The system requires only a transparent electrode on a visual screen and multiple stimulator pads attached to user's fingers for generating force feedback, and thus can be easily integrated onto a flat panel display. It presents passive (resistive) force feedback by friction force which is modulated by electrostatic attraction force between the transparent electrode and the stimulator pads. Applying different voltages to different stimulators provides independent stimulus to the multiple fingers.

The first prototype reported in [6] realized two independent haptic stimuli up to approximately 1 N on an LCD (Liquid Crystal Display) monitor. In that prototype, the positions of the fingers were detected using an over-mounted camera. Besides position detection, the system requires a force-direction sensor to solve "sticky wall" problem, which is commonly found in passive haptic systems [8], [9]. Our second prototype in [7] incorporated a simple force-direction sensor in the stimulator pads, which successfully reduced the stickiness of virtual walls.

Although those prototypes have successfully demonstrated multi-finger haptic feedback on flat panel displays, the systems relied on the overhead cameras, which tends to complicate the system structure; they should preferably be integrated with multi-touch input interfaces. Unfortunately, the indirect electrostatic stimulation is not yet compatible with typical capacitive-type touch interfaces, and thus other touch input principles need to be investigated for the integration. In this paper, we focus on the vision-based touch panel technology and investigate its compatibility with the multi-finger haptic feedback technology. Since both technologies, vision-based touch panel and multi-finger haptic technology, are designed for large-size displays, their integration is expected to realize rich user experience on a large-size tabletop display.

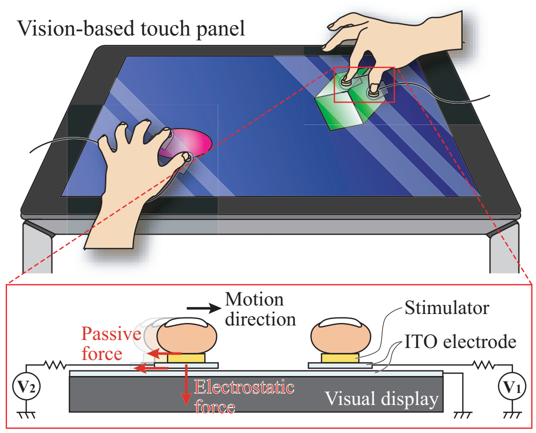
This paper is structured as follows. Section II reviews related studies on electrostatic haptic feedback. Section III explains the concept of the proposed method to realize position and force-direction detection. Section IV investigates the compatibility of the electrostatic haptic system with the vision-based touch panel, using SAMSUNG SUR40 with Microsoft PixelSense as an example. Section V discusses the design of the stimulator pad for force-direction detection. Section VI demonstrates multi-finger visuo-haptic interactions with the developed system and evaluates the performance of the force-direction detection in haptic rendering. Section VII concludes the paper.

II. Related Studies on Electrostatic Haptic Feedback

The multi-finger electrostatic haptic system described in this article utilizes indirect-type electrostatic stimulation. Electrostatic stimulation (or electrotactile stimulation/vibration), which was firstly reported in [10], utilizes friction force modulated by electrostatic force between two conductive objects facing each other. Using a transparent electrode as one of the two objects, haptic feedback can be realized on visual displays without impeding the visual information on the screen [1]–[3].

There are two types in electrostatic stimulation: direct and indirect. Direct-type system stimulates a finger by inducing electrostatic force between a finger and a transparent electrode on a display [1], [2]. It can render haptic feedback directly on the user's finger with a considerably simple setup. The system, however, requires segmented electrodes if we want to

 ISBN: 978-1-61208-325-4



Multi-finger electostatic haptic feedback system

Figure 1: System overview of proposed system

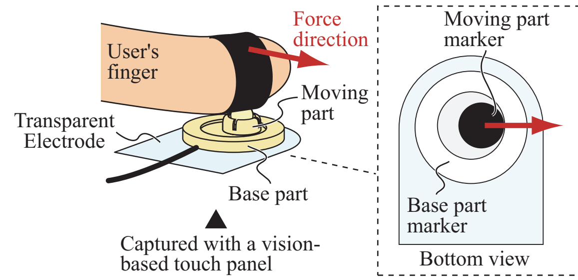
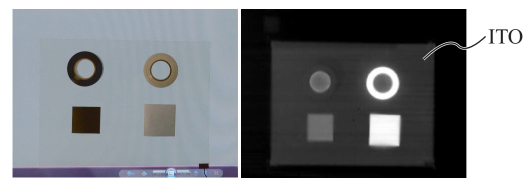


Figure 2: Concept of stimulator and marker equipped to its bottom



(a) Appearance (b) Captured image

Figure 3: Appearance (a) and captured image (b) of objects on ITO film

realize multi-touch interaction, which is practically not easy to develop. Moreover, the presented force is prone to instability due to finger perspiration [11] and can provide only weak force (approximately 0.1 N as reported in [12]). Due to the small feedback force, the system can only render surface texture sensations in cutaneous sense.

On the other hand, indirect-type system stimulates a finger through a stimulator pad, which results in larger and more stable feedback force [3]. It can be easily extended for multi-touch by employing multiple stimulator pads and applying different voltages to different stimulators [6]. One drawback of the indirect type is that the stimulator pads need to be arranged on the display, which results in more complicated system structure than the direct-type. Due to this drawback, the indirect type would not be suitable for mobile devices, such as smartphones and tablets. Indirect type is rather suitable for large-size tabletop systems, to enhance user interaction on those systems.

III. Multi-Finger Haptics on a Vision-Based Touch Panel System

A. *Prototype Overview*

Figure 1 shows the overview of the prototype system that was studied in this paper. The system integrates multi-finger electrostatic haptic feedback system as proposed in [6] on a vision-based touch panel PC (Samsung SUR40 with Microsoft PixelSense). The novelty of this work compared to the previous studies in [6], [7] is to utilize the touch panel for positions and force-directions detection, which can considerably simplify the setup. The touch panel contains IR (infrared) backlight and IR sensor array inside the LCD panel, with which the system detects the object locations on the screen. In the integrated prototype, the touch panel detects the positions of the stimulator pads, as well as applied force directions on the pads. The system then calculates feedback force based on the interaction within the visually-rendered virtual world, which is then provided to the pads such that the user can haptically feel the virtual world through their multiple fingers.

The multi-finger electrostatic haptic feedback system consists of a non-segmented transparent electrode that covers the whole surface of the display and multiple stimulator pads. A user is equipped with those pads on his/her fingertips to receive haptic stimulus. The transparent electrode is electrically grounded, and each stimulator pad is fed voltage through a voltage feeding wire. The voltage application to the pad changes vertical electrostatic attraction force between the pad and the display. The motion of the fingertip, then, converts the attraction force into horizontal friction force, which is perceived as haptic stimuli. Without fingertip motion nor putting the fingers on the pads, no feedback is given to the user. By applying different voltages to different stimulator pads, the pads can render different stimuli simultaneously.

B. *Position and Force-Direction Sensing*

Detection of position and applied force direction is realized by analyzing a captured IR image taken by the vision-based touch panel. The position means positions of stimulator pads, and the applied force direction is detected by using relative displacement of a user's finger to the stimulator pad attached to the finger. Figure 2 shows the concept of the structure of the stimulator pads that realizes the detection. Each stimulator pad consists of a base part and a moving part which the user wears on his/her fingertip. The base part is connected to a transparent electrode, to which the voltage for electrostatic stimulation is applied. The base part is ring-shaped and is marked in white on its bottom. The moving part, which is marked in black on its bottom, can be moved freely by a user within the inner circle of the base part. The system detects the position of the stimulator by calculating the center position of the white marker and force direction by comparing the black marker position against the white marker.

IV. Design Consideration

A. *Interference of ITO Film to Vision-based Touch Panel*

The applicability of ITO electrode to the vision-based touch panel using IR sensors is investigated in this section. As the ITO transparent electrode is not perfectly transparent in infrared region, its effect on the vision-based touch panel needs to be investigated. To examine the interference, some IR images were captured through a small ITO film. Figure 3 shows one of the captured images. In this image, the ITO film covers the rectangular area, on which some objects are placed. The objects are two sheets of black/white paper and two rolls

 ISBN: 978-1-61208-325-4

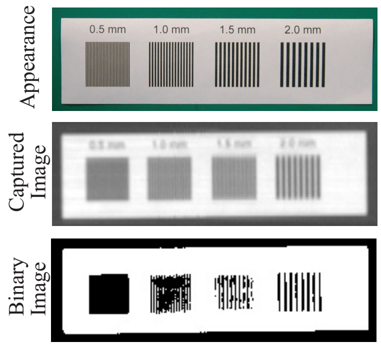


Figure 4: Appearance (top), captured image (middle), and binary image (bottom) of striped pattern

of black/white vinyl tape. The image shows that the ITO is not perfectly transparent in the IR region, but is transparent enough for object detection.

It was confirmed that the black vinyl tape completely absorbed the IR light, and thus invisible, whereas the white paper and vinyl tape were clearly visible even through the ITO film. The black paper sheet did not provide clear image, and thus is not suitable as a marker. In addition, an interesting phenomenon was observed. In the captured image, the inside of the black vinyl tape roll appears bright, although there is nothing inside. That is probably because IR light was reflected on the inside of the roll, whose color was white. It shows that the captured image is affected not only by the bottom color of the objects, but also by the vertical edge faces, suggesting that the color of edge faces should be designed carefully when making the stimulator pads.

B. Resolution of Captured Image

The resolution of the captured images through the ITO film was investigated, as the resolution is important to design marker pattern for the stimulator pads. According to the specification of the display, the display size is 885.6 × 498.15 mm and the captured image has 960 × 540 pixels, which means the pixel resolution is approximately 1 mm. However, the IR image can be blurred and practical resolution can be much worse.

To evaluate the resolution, some black striped patterns printed on white paper were observed through an ITO film. Four different stripe patterns were prepared, with different widths for black strips: 0.5, 1.0, 1.5, and 2.0 mm. The white strips had the same width as the black strips. Figure 4 shows their appearance, a captured image on the touch panel, and a binarized image. Although the boundaries of the striped patterns are blurred in the captured image, all the stripes except 0.5-mm interval were recognized as stripes even in the binary image. This pilot evaluation indicates the system can recognize 1.0 mm object through the ITO film.

C. Latency of the vision-based touch panel

The latency of the vision-based touch panel in position detection was measured. A white marker was moved laterally on the touch panel, whose position was measured by both the touch panel and a laser displacement sensor (OMRON, ZX-LD100). Figure 5 compares the measurement results. The red dashed line indicates the reference position measured with the laser sensor. The blue line indicates the position detected by the touch panel. The latency of the touch panel in position detection was found to be about 80 ms.

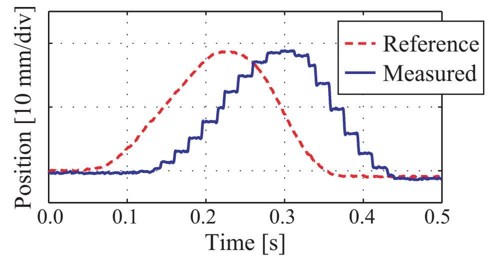


Figure 5: Latency of the vision-based touch panel in position detection.

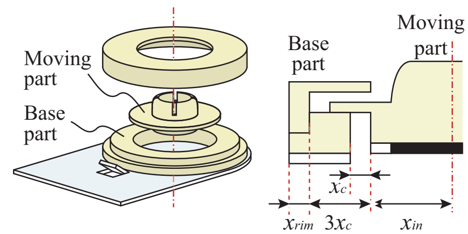


Figure 6: Assembly diagram (left) and cross section (right) of stimulator pad

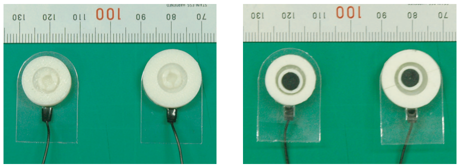
The latency limits the maximum motion speed for proper operation on the electrostatic haptic system. If we suppose 4-mm tolerance for position detection, the maximum speed is limited to 0.05 m/s for the latency of 80 ms. The speed is much smaller than a desired speed: e.g. 0.5 m/s, which was realized in the previous tracking system in [7] using an overhead motion-tracking camera, which has 8-ms latency. The desired speed could be achieved by compensating the latency with some kind of predictive control, which would be implemented in our future work.

V. Force-Direction Detecting

A. Design of Stimulator Pad and Its Marker

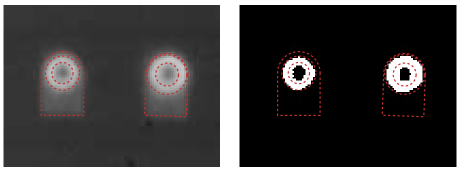
Figure 6 shows the assembly diagram and cross sectional view of the proposed stimulator pad to facilitate force-direction detection, as well as position detection. The stimulator pad consists of a base part fixed on a small ITO electrode and a moving part fixed to a user's finger. The base part has a cap to prevent the moving part from popping out from the base part. The proposed system calculates the applied force direction by detecting relative displacement of the moving part to the base part. Thus, it requires large clearance for accurate detection of force-direction. According to the resolution of captured image, the clearance must be larger than 1.0 mm. The accuracy of force-direction detection is expected to be higher as the clearance becomes larger. However, large clearance enlarges the size of stimulator pad, in trade-off. As shown in Figure 6, the radius of the stimulator r becomes $r = x_{rim} + 3x_c + x_{in}$, where x_{rim} is the width of the cap part, x_c is the clearance between the base and moving part, and x_{in} is the radius of the marker for the moving part.

We determined the size so that the width of the stimulator pad becomes comparable to the width of human finger, to avoid occlusion of visual information. Figure 7 shows the appearance of prototype stimulator pads. Two stimulator pads with different design parameters were prepared for evaluation: the smaller one has 1.0 mm clearance and its diameter is 16



Top view Bottom view

Figure 7: Appearance of stimulator. Top view (left) and bottom view (right)



Captured image Binary image

Figure 8: Captured image (left) and binary image (right) of prototype stimulator on the vision-based touch panel

mm, and the larger one has 1.5 mm clearance and its diameter is 19 mm. They have common width of the cap (1 mm) and radius of the moving part marker (4 mm). To obtain high contrast images at the border between the black and the white marker, a black circle marker (ϕ6 mm) enclosed by white ring (1 mm width) is attached to the bottom of the moving part.

Figure 8 shows a captured IR image and a binarized one for the two stimulator pads placed on the touch panel covered with the ITO film. Red dashed lines in the figure indicate the contours of the stimulator electrodes and those of the base-part markers. The clearance between the base part and the moving part is blurred in the captured image and not clearly visible. Yet, the black marker on the moving part is clearly observed as a black circle, and thus it is expected that the force direction can be calculated from the relative position between the black circle and surrounding white circle, for both prototype pads.

B. Evaluation

The force-direction detection capability was evaluated for the two prototypes in a static condition. The stimulators were fixed on the touch panel displaying no visual image, and their moving part was pushed to 8 directions in every 45°. The force-direction (which is pushed direction) was calculated from the relative position of the markers. Then, the mean, maximum, and minimum values of detected direction for two seconds in every direction were measured. Figure 9 and Figure 10 show the results for the smaller prototype and the larger prototype, respectively. Both prototypes succeeded in detection, but their resolutions are considerably different. The smaller prototype showed an error up to approximately 45°, while the larger prototype indicated an error up to approximately 15°. It confirms that the larger clearance increases the accuracy of the force-direction detecting, as expected.

Although the above static evaluation was conducted on the black screen, if there is some visual image displayed on the screen, it can influence the performance of force-direction detection. An example of such influence is shown in Figure 11. When the stimulator pad is placed on the border of black and white image as shown in Figure 11(a), the touch panel captures a distorted image of the marker as the images on the display is reflected by the ITO film and the makers, as in Figure 11(b). In this case, the system cannot detect the force-direction correctly,

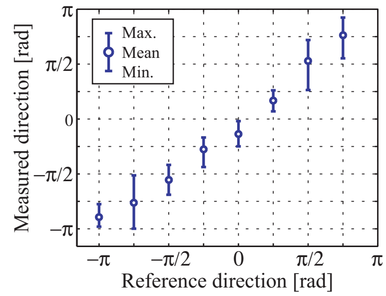


Figure 9: Results of force-direction detecting of small clearance pad

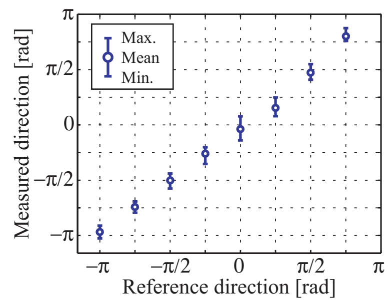
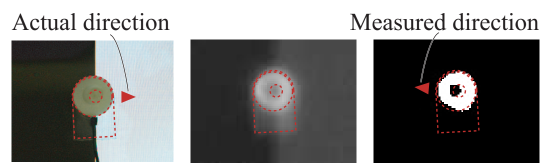


Figure 10: Results of force-direction detecting of large clearance pad



(a) (b) (c)

Figure 11: Failure case of force-direction detecting. Appearance (a), captured image (b), and binary image (c) of stimulator at the border on black and white background.

as shown in Figure 11(c) that shows completely opposite force direction.

There are two possible approaches to solve this problem. One is to fix the image under the stimulator. By tracking the stimulator position and drawing a constant-colored background below the stimulator, the system can fix the condition for force-direction detection. This approach, however, limits the motion speed of the stimulator pad quite low, since the position tracking speed of the touch panel is not so fast in this system.

The other is to render virtual environments in low contrast colors. Using colors which are captured in similar brightness, such like black and blue, the system can reduce the marker distortion, even when the stimulator pad is placed on a border between different colors. This approach limits coloration of virtual environments, but it is not affected by the tracking speed. In the final experiment described in section VI-B, we adopted this latter approach.

To evaluate the dynamic performance, the response time was measured with the large-clearance stimulator. The response time, or latency, of the force-direction detection was found to be about 80 ms, which is almost the same as the latency of the touch panel.

 ISBN: 978-1-61208-325-4

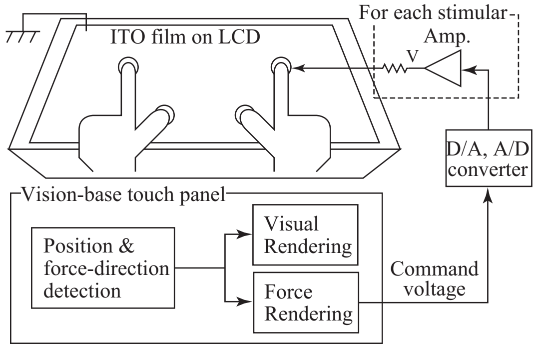
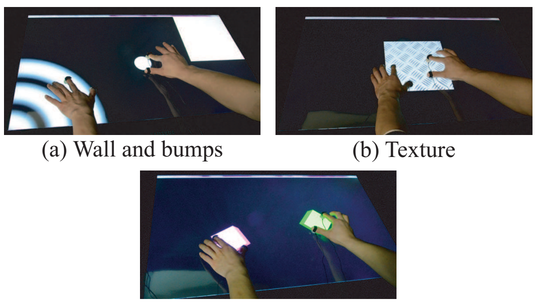


Figure 12: Prototype setup



(a) Wall and bumps (b) Texture

(c) Dynamically moving cubes

Figure 13: Examples of multi-finger visuo-haptic interaction

VI. Pilot Experiment for Haptic Rendering

A. Demonstrations without force-direction detection

Figure 12 shows the setup of the whole experimental system. The system consists of the SUR40, whose surface is covered with a transparent ITO electrode, four stimulator pads, a DA/AD converter board, and four high-voltage amplifiers. The ITO electrode was grounded electrically, and its surface was insulated with a PET film (8 μm in thickness). The four stimulator pads are the larger one of the two types discussed. The bottom surface of stimulator ITO electrode was also covered with the PET film.

In [6] and [7], rendering of static objects, such as walls, bumps, and surface textures, as well as dynamically moving objects, were reported. First, the same interactions have been implemented, but without force-direction detection, on the developed system as shown in Figure 13. In these interactions, the rendering program runs at approximately 80 Hz. When the system did not employ the force-direction detection, if we ignore the sticky wall and object stiction [7] problems, the system could successfully render the haptic reaction force, although the maximum operation speed of the contact pads was limited to several centimeters per second for proper operation. This limitation was due to the low sampling rate and the latency of the touch panel.

B. Evaluation of Force-Direction Detection in Wall Rendering

In a passive haptic rendering, a surface of a static virtual wall can be felt sticky. This stickiness comes from the fact that passive friction force always acts toward the opposite direction of user's motion. Typically, a virtual wall is described as a spring; if a contact point is penetrating the surface, a force proportional to the penetration depth is fed back, which should be directed toward outward direction. When the contact pad is proceeding into the wall, there is no problem as the reaction force is in the opposite direction of the moving direction. On the other hand, when retreating, the force direction and the moving direction are the same, which is not possible to realize on a passive system.

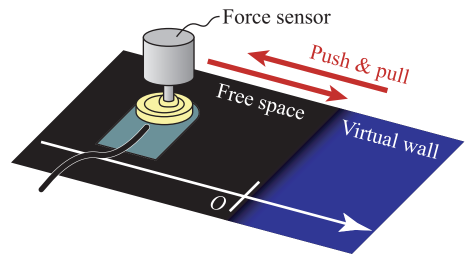


Figure 14: Setup of virtual wall rendering evaluation

Typical solution for this is to cut the force rendering during retreating. To facilitate that, the system needs to detect in which direction the user is trying to move the contact pad. In the developed system, the direction of the applied force is achieved by the force-direction detection mentioned above. Thus, the following evaluates if the force-direction detection method can successfully eliminate the sticky wall problem.

The schematic illustration of the experimental setup dedicated for the purpose is shown in Figure 14. In this setup, only one pad was used whose motion is limited in one-dimension by using a linear guide. The stimulator was manually moved to enter into and retreat from a virtual wall, during which the rendered force was measured with a force sensor (Nitta, PD3-32). Virtual wall was visually rendered using blue color on black background to prevent the malfunctioning of the force detection, as mentioned before.

Figure 15 shows the result of virtual wall rendering, first without considering force direction. The plot on the top shows the center position (solid line in the top) and the both edges (dashed line in the top) of the stimulator pad, calculated from the IR captured image. The penetration depth was calculated using the edge position. The other two plots are for the applied voltage and the measured force. When the stimulator pad hits and pushed on the wall (the first half of the plots), the reaction force from the wall was rendered correctly. However, in the latter half, force is exerted in negative direction while the operator was trying to retreat from the surface. This means that the contact pad is sticking to the virtual wall.

Figure 16 shows the result of virtual wall rendering that considered applied force direction. The former half is almost the same as the previous experiment, but the latter half shows the distinct difference. As the force direction was altered, the applied voltage was cut to zero, which allowed the stimulation pad to retreat from the virtual wall easily. Although, some negative force was observed, its magnitude and duration was considerably reduced, which confirmed the effectiveness of the proposed system.

C. Discussion

First of all, the most significant problem in the current system is the limited operation speed. In all the experiments, the operation speed of the stimulation pads was limited up to several centimeters per seconds for proper operations. This is due to the limited tracking performance of the touch panel.

 ISBN: 978-1-61208-325-4

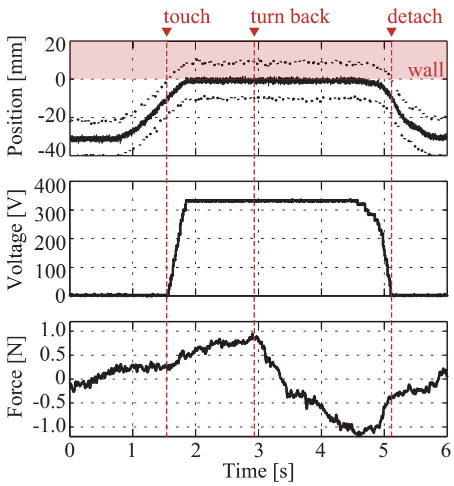


Figure 15: Result of virtual wall rendering without force-direction

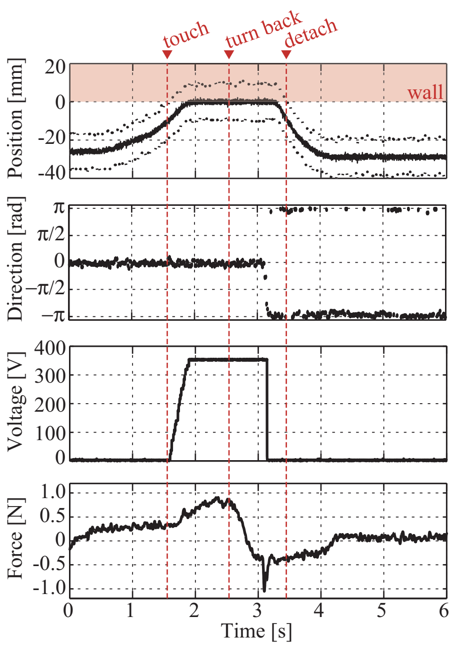


Figure 16: Result of virtual wall rendering with force-direction

For force-direction detection on a virtual wall, two major problems were found. One is vibration during detaching (which is not observed in Figure 16). As the force-direction detection is not robust enough, it sometimes outputs wrong direction. If the detection malfunctions during detaching from a wall, it results in rapid on and off of the voltage that creates vibrative sensation. This would be solved by having a dead zone in force-direction detection and to use low pass filtering in the detection result.

The other problem was the negative spike in rendering force when the user changed the force direction (which is clearly observed in Figure 16). This negative spike would be the result of the latency of the information processing. As there is time-lag between the real motion and the image capturing, the voltage was not instantly switched off as the user alters the force direction. If a user rapidly alters the force direction, the time-lag of voltage-off results in the spike of the rendered force.

The first and the third problems are both due to the processing speed of the touch panel. The problems might be solved by implementing some kind of predictive control, which we would like to work on in our future study.

VII. Conclusion

This article described position and force-direction detection using a vision-based touch panel, for multi-finger electrostatic surface haptic system. Using the embedded IR capturing system in the touch panel, the system can detect the positions and applied force directions of the stimulator pads, without relying on an overhead camera.

The paper investigated compatibility of the vision-based touch panel with the electrostatic stimulation, based on which a prototype system was designed and demonstrated. The prototype realized various visuo-haptic interactions, but the motion speed was limited due to the slow tracking speed of the touch panel. The system demonstrated successful force-direction detection, but at the same time, some limitations have become clear, which are the vulnerability against visual images and slow detection speed due to the slow tracking speed.

Acknowledgment

This work was supported by Funding Program for Next-Generation World Leading Researchers (LR013) from Japan Society for the Promotion of Science.

References

[1] O. Bau, I. Poupyrev, A. Israr, and C. Harrison, "Teslatouch: Electrovibration for touch surface," in Proc. 23rd annual ACM symposium on User Interface Software and Technology, 2010, pp. 283–292.

[2] J. Linjama and V. Mäkinen, "E-sense screen: Novel hapitc display with capasitive electrosensory interface," in HAID 2009, 4th Workshop for Haptic and Audio Interaction Design, 2009.

[3] A. Yamamoto, S. Nagasawa, H. Yamamoto, and T. Higuchi, "Electrostatic tactile display with thin film slider and its application to tactile telepresentation systems," IEEE Transactions on Visualization and Computer Graphics, vol. 12, 2006, pp. 168–177.

[4] S. Saga and K. Deguchi, "Lateral-force-based 2.5-dimensional tactile display for touch screen," in Proc. 2012 IEEE Haptic Symposium, 2012, pp. 15–22.

[5] M. Weiss, C. Wacharamanotham, S. Voelker, and J. Borchers, "Fingerflux: near-surface haptic feedback on tabletops," in Proc. the 24th annual ACM symposium on User interface software and technology. ACM, 2011, pp. 615–620.

[6] T. Nakamura and A. Yamamoto, "Multi-finger electrostatic passive haptic feedback on a visual display," in Proc. IEEE World Haptics Conference 2013, 2013, pp. 37–42.

[7] ——, "Multi-finger surface visuo-haptic rendering using electrostatic stimulation with force-direction sensing gloves," in Proc. 2014 IEEE Haptics Symposium, 2014, (accepted).

[8] J. E. Colgate, M. A. Peshkin, and W. Wannasuphoprasit, "Nonholonomic haptic display," in Proc. 1996 IEEE International Conference on Rotobitcs and Automation, 1996, pp. 539–544.

[9] J. Furusho, M. Sakaguchi, N. Takesue, and K. Koyanagi, "Development of ER brake and its application to passive force display," Journal of Intelligent Material Systems and Structure, vol. 13, 2002, pp. 425–429.

[10] E. Mallinckrodt, A. L. Hughes, and W. S. Jr., "Perception by the skin of electrically induced vibrations," Science, vol. 118, 1953, pp. 277–278.

[11] H. Tang and D. J. Beebe, "A microfabricated electrostatic haptic display for persons with visual impairments," IEEE Transactions on Rehabilitation Engineering, vol. 6, 1998, pp. 241–248.

[12] D. J. Meyer, M. A. Peshkin, and J. E. Colgate, "Fingertip friction modulation due to electrostatic attraction," in IEEE World Haptics Conference 2013, 2013, pp. 43–48.

TouchPair : Dynamic Analog-Digital Object Pairing for Tangible Interaction using 3D Point Cloud Data

Unseok Lee and Jiro Tanaka
Department of Computer Science
University of Tsukuba
Tsukuba, Ibaraki, Japan
{leeunseok, jiro}@iplab.cs.tsukuba.ac.jp

***Abstract*— Sensor-based pairing technology between digital objects for interactions are used widely (e.g., smart phone to Bluetooth headset). In addition, research about interactions between daily normal analog objects (e.g., a doll, Lego block) and digital objects has progressed and is also popular. However, such research can only involve interactions with pre-setup objects. The paired objects cannot be changed dynamically. In this paper, we propose a new analog-digital object pairing method by intuitive touch interactions using three-dimensional point cloud data. Several touch pairing methods are described in detail and paired objects are changed dynamically using the proposed method. In addition, a simple tangible interaction between two objects is described after pairing. Finally, we demonstrate the high recognition rate of the proposed method using experiments and describe our system's contribution.**

Keywords-dynamic pairing; point cloud; tangible interaction; 3d gesture;human computer interaction

I. INTRODUCTION

In everyday life, the touching action is natural and common. We touch objects to use them (e.g., a doll or toy to play, open a bottle cap for drinking). Touch interactions with digital devices have also become natural in recent years, because smart devices with touch screens and touch pads are now used widely.

Figure 1. Analog-Digital objects in everyday life

Simultaneously, in the field of human–computer interaction (HCI), research on interactions between physical objects and digital devices has progressed rapidly. A physical object is set as an input unit and the digital device is controlled by it. Such interactions are used widely and have become a 'natural' method. However, to use a physical object as an input unit, much effort and time is initially needed to set up sensors. Moreover, it takes time and effort to apply sensors again when changing the physical object as the input unit. In addition, the digital object is limited to a particular physical object. Thus, there is no 'natural' interaction between various objects. Regarding the input unit, research on methods for making a tangible object for which touch sensing is possible has progressed. For example, in the bowl project [2] , a simple media player in a bowl sits on a living room table and a range of physical objects can be placed within it. When an object is placed in the bowl, related media are played on the TV. The project used radio-frequency identification (RFID) sensors for tangible interactions. However, the system could not provide dynamic pairing between objects. The interactions and possible objects were also limited.

The "HandSense" [3] prototype used capacitive sensors for detecting when it was touched or held against a body part. It could determine whether a device was held in the left or right hand by measuring the capacitance on each side. Raphael [4] presented a method for prototyping grasp-sensitive surfaces using optical fibers. However, all of these examples require attaching sensors to the devices. This is unnatural in the real world. Also, the paired object cannot be changed dynamically.

In this paper, we propose a new method for dynamic pairing with tangible interactions between analog objects and digital objects in practical circumstances. This dynamic pairing is achieved through touch interactions. For example, one hand grasps a doll, an analog object. Then, the other hand grasps a smart phone, a digital object. The doll and smart phone are paired and prepared for tangible interactions. The system makes it possible to pair the doll with touch in three dimensions. The smart phone then shows feedback from interactions with the doll. We can change the paired objects dynamically. Figure 1 shows examples of pairing analog and digital objects in everyday life. Our pairing technique is based on three-dimensional(3D) point cloud data using two Kinect units. They capture and calibrate 3D point cloud data. Our system determines touch pairing and tangible interactions of the paired analog object, based on these calibrated data. In this way, the system can readily recognize what objects are touched and trace what objects are paired. In addition, we can recognize the touched position and movements of the objects. We present the results of tests of recognition rate for pairing using the proposed method.

The rest of this paper is organized as follows. Section 2 introduces related work on depth-based touch sensing and tangible interactions. We describe in detail the principles of the pairing method, the system specifications, and tangible interactions in section 3. We present details on the high recognition rate of our system in section 4. Finally, we describe our contribution and future work in section 5.

II. RELATED WORK

A. Depth-based Touch Sensing Technologies

In recent years, depth-based cameras and related technology have developed rapidly. Research on obtaining 3D data on objects using depth information has also progressed. The framework for 3D sensing using depth cameras has been improved markedly [5].

Florian et al. implemented tangible interactions using a depth camera and a 3D sensing framework [1]. They implemented touch detection and object interaction, supporting multi-touch and tangible interactions with arbitrary objects. They used images from a depth camera to determine whether a user's finger touched the object. However, they were unable to support 3D touching and dynamic pairing between objects for tangible interactions.

Andrew et al. used depth-sensing cameras to detect touch on a tabletop [7], using the camera to compare the current input depth image against a model of the touch surface. The interactive surface need not be instrumented in advance for the interaction and this approach allows touch sensing on non-flat surfaces. However, they only supported simple touch recognition and could not address touch in any direction with 3D objects.

B. Tangible Interactions with Analog Objects

"Digital Desktop" by Wellner et al. [8] was used in an early attempt to merge the physical and digital worlds. They implemented a digital working space on a physical desktop where physical paper served as an electronic document. The interaction with papers was by means of bare fingers. "Icon Sticker" [9], based on this idea, is similar. Icon Sticker is a paper representation of digital content. It consists of transferring icons from the computer screen to paper, so they can be handled in the real world and used to access digital content directly. An icon is first converted into a corresponding barcode, which is printed on a sticker. Then the sticker can be attached to a physical object. To access the icon, the user scans the barcode on the sticker with a barcode scanner. "Web Sticker" [10] uses barcodes to represent online information. It is similar to Icon Sticker, but instead of icons it manages Web bookmarks. They use a handheld device with a barcode-reading function to capture the input and display related information.

There were also attempts to improve tagging of physical objects for a more natural tangible interaction. Nishi et al. [11] registered real objects on a user's desktop based on a user indicating a region on the desk by making a snapshot gesture with four fingers. A color histogram was used to model the object and a pointing gesture was used to trigger the recognition. "Enhance Table" also uses a color histogram to model objects. However, the size is predefined and the system is limited to mobile phone recognition.

Although many previous tangible interaction studies have used physical objects for interactions, most of them are token-based approaches and provide only limited use of real objects. They do not support 3D object tracking or pairing for tangible interactions. Thus, to overcome this, we propose a robust 3D object-tracking method that detects touch in three dimensions. The system supports dynamic pairing between analog and digital objects, and makes analog objects accessible to touch anywhere.

III. TOUCHPAIR

A. Hardware and Software

Our system consists of two Microsoft Kinect sensors for Xbox 360 with stands. Two computers (Intel i7 2.4 Hz, 8 Gb RAM, and GeForce GTX 660M graphics card) are used to handle the 3D data. A desk and the pairing object for interaction are installed. The analog and digital objects are placed anywhere. The Kinect sensors are set at 80 cm from the desk. The computers are connected to each Kinect sensor, and the digital objects have wireless internet or Bluetooth connections with the computer, installed in the bottom of the desk (Fig. 2).

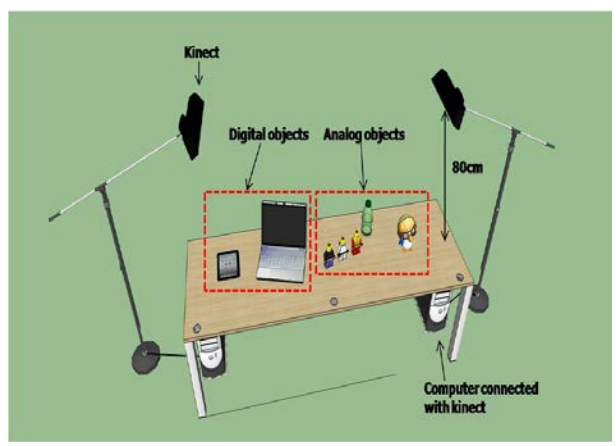


Figure 2. TouchPair System Configuration

Our system uses OpenFrameworks OpenNI [6] for the Kinect sensors and a point cloud method that provides example add-ons of frameworks. The system obtains 3D point cloud data and maps the RGB data to the point cloud. The movement of the points is based on pairing recognition. The proposed system was implemented on a Microsoft Windows 7 platform. The pairing recognition module was implemented in Visual Studio 2010 and OpenNI 1.5.4.

B. TouchPair Architecture

The entire system consists of three major modules(Fig. 3). The input data are obtained by the two Kinect sensors, on the left and right sides. The data are sent for processing. The process is detailed below.

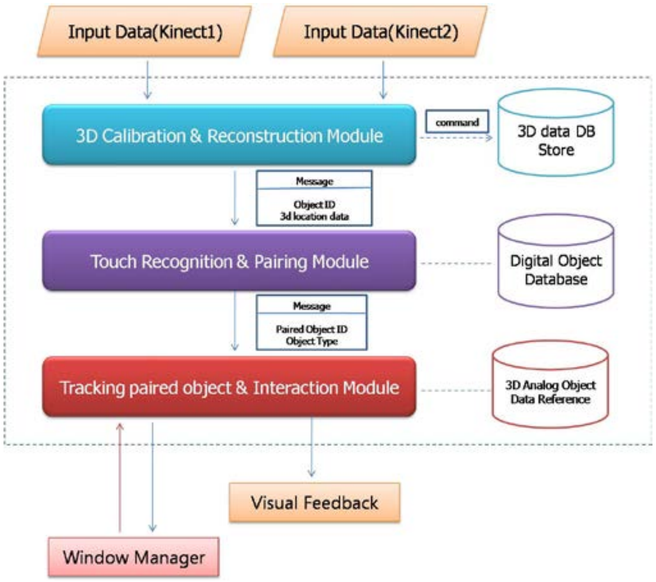


Figure 3. Diagram of TouchPair System Architecture

1) 3D calibration and reconstruction module: In this module, we calibrate depth data for each object, obtained from the two Kinect sensors. The system makes a 3D reconstruction using a point cloud library with calibrated data. The module commands store calibrated and reconstructed data in a database, which is then used by the touch-recognition module. After storage, the module sends messages to the touch-recognition and pairing module about object ID and object location using the 3D point cloud data.

2) Touch-recognition and pairing module: Touch is recognized in terms of the depth and position of the object and hands using 3D tracking. Using the previous depth information from the 3D reconstruction based on the point cloud, the system determines whether the hand touched the object, and if so, the position of the object. The system recognizes the time of touching between the user's 3D hand point cloud data and the object 3D point cloud data, then determines whether they are paired. A paired analog object's 3D point cloud data are stored and sent to the tracking module with information on the object type. We defined a limited objects database.

3) Module for tracking paired objects and interactions: After pairing, the system tracks the analog object based on saved 3D point cloud data. The paired digital object can be tracked; however, the paired objects do not commonly move. The paired analog object is tangible, based on 3D analog object data, from the 3D calibration and reconstruction module. We can make an interaction with the digital object in this module; the interaction is shown by visual feedback.

C. TouchPair Method using 3D Point Cloud Data

Our proposed method uses 3D point cloud processing of Kinect depth data. A point cloud itself is a set of data points in a coordinate system. We measured a large number of points on the surface of an object using OpenFramework [6].

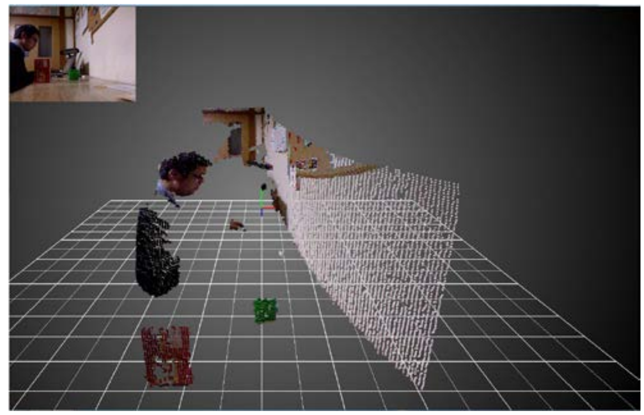


Figure 4. 3D Point Cloud Data

The system obtains RGB data from two Kinect sensors (Fig. 4) and assigns them to the depth area. However, all directions of the object cannot be reconstructed. Thus, we find the most appropriate location for the Kinects and position them so that they cover most of the experimental space.

1) Touch gesutre and recognition: Our touch pair system recognizes the touch actions of users' hands based on depth. The system calculates the depth of each point between a user's hand and the object by filtering closer data. The flow of recognition is as follows:

a) Calculation of all point depths: Calculate all points of the analog and digital objects and the user's hand on the table.

b) Determination of finger position: The system calculates the minimum and maximum depth of the finger by defined thresholds because we hold our fingers in specific ways when we touch something.

c) Determination of hand position: The system calculates the minimum and maximum depths of all fingers and the palm; from the front view, the system uses depth information from both sensors simultaneously.

d) Determination of grasping: Using depth data on users' hands and on objects collected from both sensors simultaneously, we found certain threshold values for recognizing the act of grasping.

2) Analog-digital object pairing: Our system performs time calculations between touched analog-digital objects versus touched object-object pairing. Implementation of object-object touching is shown in Figure 5c,d. When the system recognizes that the user wants to pair, the color is changed. The red color refers to the digital object and the analog object is blue. The recognized hand is shown in yellow using 3D point cloud data. The main steps are as follows.

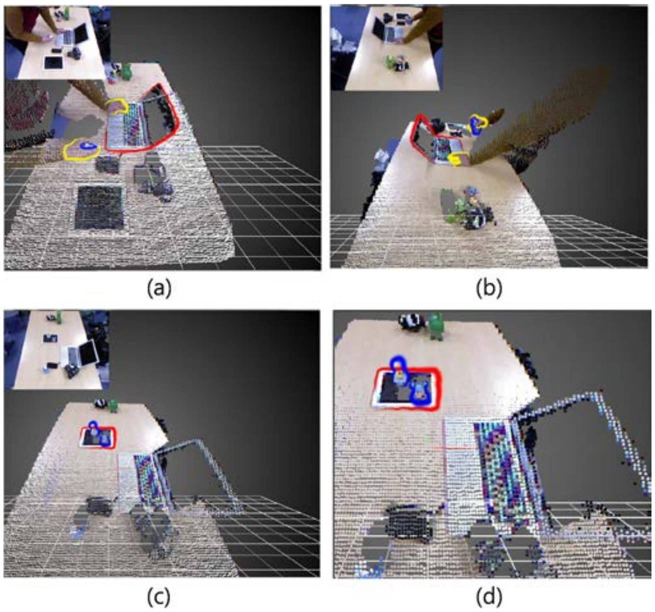


Figure 5. (a) Pairing gesture with hand touching to digital object and grasping analog object from right camera (b) Pairing gesture from left camera (c) Pairing by object-object touching, toy and smart pad are pairing (d) Detailed object-object pairing image

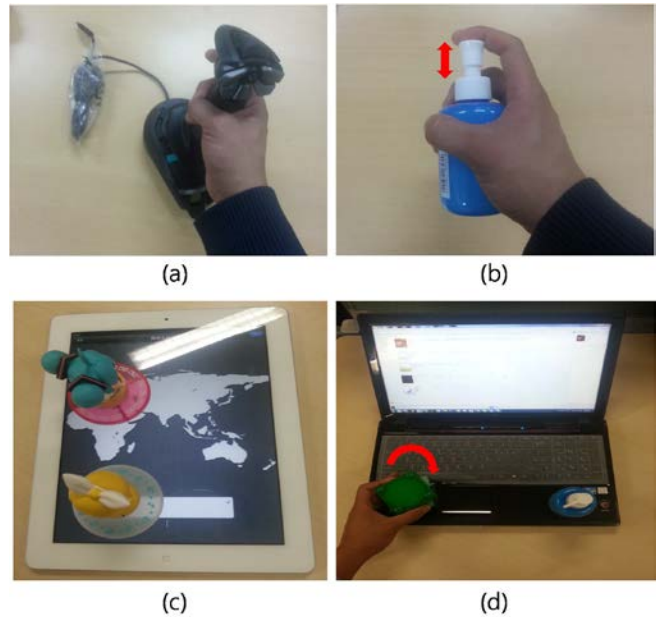


Figure 6. (a) Flight joystic controller (b) Paired analog object controller as flight joystic (c) Control the map by paired two toys (d) Control media player by paired green cube by rotation.

a) Time calculation: For pairing, the user maintains a touching posture for a few seconds after touching is recognized between the objects. The color is changed after the pairing.

b) Tracking a paired set: To track paired objects, the system calculates 3D point cloud data continuously, which are provided by the real-time reconstruction module (Fig. 3). The color of the tracked object is shown.

c) Changing a paired set: To change a paired object set, a pairing gesture is made for some time period. The user touches what he/she wants to pair. After a few seconds, performing the pairing gesture (Fig. 5) will change the pair set as indicated by the color feedback.

D. *Tangible Interactions*

After pairing objects, the system can be used in various tangible interactions. We designed a flight joystick using analog objects. Such a joystick can be used when playing a flight or gun game. Our system made a pairing between a notebook PC and a lotion bottle with a cap as a game controller. The game can be controlled by pushing the lotion bottle's cap (see Fig. 6b) and moving it. The movements can be recognized based on the cloud data for the object and the hand. We also designed a map controller with paired objects (see Figs. 5d and 6c). Our system tracks the two toys and the touched position; the map is moved to provide feedback.

Finally, we implemented a Windows Media Player controller with a paired green cube (see Fig. 6d). The system recognizes rotation of the cube's cap, then changes the volume as feedback. The application changes songs when the lotion bottle cap is pressed.

IV. EVALUATION

We evaluated the touch pair system focusing on recognition accuracy. We evaluated touch recognition while changing the analog and digital objects alternately. We also evaluated 3D object recognition and tracking to demonstrate the system's usability and robustness.

A. *Recognition Accuracy Experiment*

Four analog objects and three digital objects were used. We evaluated finger touching, hand touching, grasping, and object-object touching for each object. The experiments were performed on computers (Intel Core i7 CPU, 2.5 GHz, and 8.0 Gb RAM) using two Microsoft Kinect sensors for Xbox.

We performed the experiments with 10 volunteers. We explained each touch pairing method sufficiently. Then a volunteer performed four touching gestures to each object 100 times. We defined 3 s as the period for completing pairing via touching. When the system recognized a pairing, it showed color feedback. Touch recognition success alone was not counted. The participants were allowed to touch analog objects only with the finger and digital objects only with the hand. Table 1 shows the average pairing recognition success for the 10 volunteers.

Our proposed system showed >90% pairing recognition with the objects provided. We found that the average finger-based pairing recognition rate was higher than hand-based pairing. Finger pairing was recognized best when one or two fingers were used. Hand pairing requires checking whether the palm is touching. Thus, hand-based pairing recognition was less accurate than finger-based pairing. In addition, the success rate for touching a smart phone was lower than that for touching the other objects. This may have been because of the size of the object. Most adult hands are bigger than

 ISBN: 978-1-61208-325-4

most smart phones. Thus, it becomes difficult for the system to find the positions of the finger and palm.

TABLE I. FINGER AND HAND TOUCH PAIRING RECOGNITION RESULT

Digital / Analog	Note PC		Smart Pad		Smart Phone	
	finger	*hand*	*finger*	*hand*	*finger*	*hand*
Toy	98.3	95.4	99.1	95.3	95.3	93.2
Black Doll	95.7	93.3	96.2	94	92.2	90.1
Green Cube	98.4	96.7	99.3	95.7	96	94
Pet Bottle	96.7	95.3	97.1	95.4	93	91

percentage of recognition rate(%)

Table 2 shows the results for other pairing methods, such as grasping and object-object pairing recognition. The experiments were performed in the same way as described in Table 1.

TABLE II. GRASPING AND OBJECT-OBJECT TOUCH PAIRING RECOGNITION RESULT

Digital / Analog	Note PC		Smart Pad		Smart Phone	
	grasp	*object*	*grasp*	*object*	*grasp*	*object*
Toy	91.4	94.3	89.1	98.1	87.3	94.2
Black Doll	85.4	95.4	85.1	97	85.2	91.1
Green Cube	91.2	96.2	90.1	96.7	87	96
Pet Bottle	90.1	97.1	90.2	97.4	88	94.3

percentage of recognition rate(%)

Grasping-based pairing recognition accuracy was >85% with the objects provided. This method uses point data from many directions. Generally, the front, side, and back surfaces of an object are touched when holding something with the hand. Thus, grasping has to be determined by analyzing the data from many directions. Thus, on the whole, the recognition rate was low. We also found that the recognition rate differed by object size and hardness. A plastic toy, green cube, and bottle are relatively hard. However, a doll is very soft. When the user touches or grasps a softer object, the system has difficulty determining touch depth. Thus, recognition accuracy was lowest for the doll among all objects provided. However, generally, the recognition rate was high.

B. Real-time 3D Object Tracking Accuracy

We evaluated object tracking after pairing. We moved analog objects during a 10 min experimental period (e.g., left-right, front-back).

Figure 7 shows the recognition accuracy for these tests. We obtained recognition rates of >80% for all objects. Because we used two Kinect sensors for real-time 3D object reconstruction and data comparisons, we obtained low error rates.

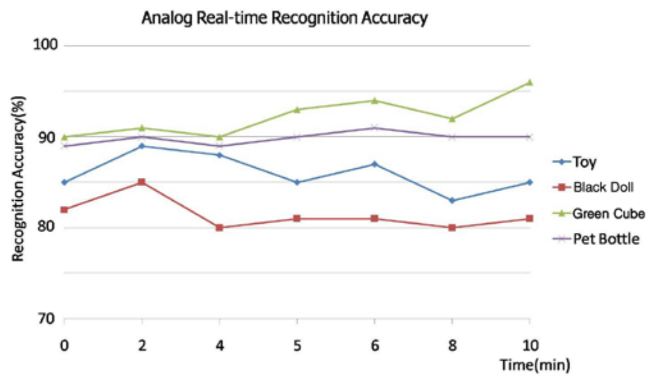


Figure 7. Analog Objects Recognition Accuracy

Figure 8 shows the recognition accuracy for three digital objects over 10 min. Smart phone recognition was <80% in around 4 min and 8 min after tracking. This is because the user was holding the smart phone with his/her hand. In particular, when we moved the smart phone with a front-back motion, the recognition rate decreased.

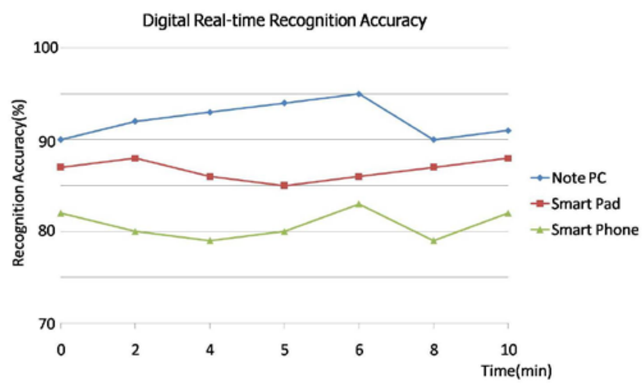


Figure 8. Digital Objects Recognition Accuracy

When the smart phone was near the user's body, the errors likely occurred because the smart phone "disappears" from the camera. However, we can overcome this problem by incorporating 3D data learning (obtained experimentally) into our system.

V. CONCLUSIONS AND FUTURE WORK

Our new dynamic analog-digital pairing method uses 3D point cloud data to assess tangible interactions dynamically. We used dynamic pairing based on finger and hand touching, grasping, and object-object touching. The accuracies were >90% for finger and hand touching and >85% for grasping and object-object touching. Almost the same results were obtained when we changed the locations of pairs dynamically. We obtained good real-time 3D object tracking results as well, despite using objects of different size, shape, and hardness.

The contributions of the present study can be summarized as follows. First, we have provided a new pairing method. Using this method to dynamically pair analog and digital objects, various tangible interactions can be achieved. Second, there is no additional device or sensor, so varied object recognition and pairing are possible. Third,

 ISBN: 978-1-61208-325-4

we obtained good recognition rates and tracking results using the proposed method.

In future work, we expect to pair users and digital objects using this system. After pairing a user and a digital device, a digital object can be controlled by the user's gestures and motions. In addition, we expect to develop various tangible interaction applications based on our proposed system. For example, an augmented graphical support interface with a paired object using a projector or head-mounted display could be developed.

References

[1] F. Klompmaker, K. Nebe, and A. Fast, "dSensingNI: a framework for advanced tangible interaction using a depth camera," TEI '12 Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction, pp, 217-224.

[2] E. S. Martinusseon, J. Knutsen, and T. Arnall, "Bowl: token-based media for children," DUX '07 Proceedings of the 2007 conference on Designing for User eXperiences, Nov. 2007, Article No. 17.

[3] R. Wimmer and S. Boring, "HandSense: discriminating different ways of grasping and holding a tangible user interace," in TEI '09 Proceedings of the 3rd International Conference on Tangible and Embedded Interaction, pp. 359–362.

[4] R. Wimmer, "FlyEye: Grasp-Sensitive Surfaces Using Optical Fiber," TEI '10 Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction, pp.245–248.

[5] OpenNI 3D sensing framework : Last visited on Nov, 2013. http://www.openni.org.

[6] openFrameworks : Last visited on Aug, 2013. http://www.openframeworks.cc.

[7] A. D. Wilson, "Using a depth camera as a touch sensor," ITS '10 ACM International Conference on Interactive Tabletops and Surfaces, pp 69-72.

[8] P. Wellner, "The digitaldesk calculator: Tangible manipulation on a desk top display," in 4th annual ACM symposium on User interface software and technology UIST 1991, pp.27-33.

[9] I. Siio and Y. Mima, "Iconstickers: converting computer icons into real paper icon," in 8th International Conference on Human-Computer Interaction: Ergonomics and User Interfaces, pp. 271-275.

[10] P. Ljungstrand, J. Redstrom, and L. E. Holmquist, "Webstickers: using physical tokens to access, manage and share bookmarks to the web," in Designingaugmented reality environments DARE 2000, pp.23-31.

[11] T. Nishi, Y. Sato, and H. Koike, "Interactive object registration and recognition for augmented desk interface," in IFIP conference on human-computer interface Interact 2001, pp.240-246.

Rapid Prototyping Spiral for Creative Problem Solving in Developing Countries

Shohei Aoki, Koichi Hori
School of Engineering
The University of Tokyo
Tokyo, Japan
emails: {aoki@ailab.t.u-tokyo.ac.jp, hori@computer.org}

***Abstract*—In this paper, we focus on how we can assist the local designers in the developing countries to design and manufacture the problem solving products. Although there are a lot of examples of appropriate technology, most of them were designed by the experts who have professional knowledge. To solve this issue, we propose a design method *rapid prototyping spiral*, considering the conventional studies on creativity support. The rapid prototyping spiral is a quick iterating process of creating prototypes and getting feedbacks from the users. The advantage of this method is that it enables the designers to discover the hidden issues or the unconscious assumptions underlying the users or the community. Our method also takes advantage of the limitation of materials, which has been regarded as a negative aspect so far. We illustrate how our method is applicable to design in the developing countries, taking a project in Ghana as an example.**

Keywords-Design method; Prototyping; Design space.

I. INTRODUCTION

A. Problem solving products in developing countries

Creating the innovative product to solve daily problems gives great impact especially to the people in the developing countries, where they are exposed to severe constraints. One of the famous activities to make these products is the *appropriate technology*. The appropriate technology is an application that is small-scale, labor-intensive, energy-efficient, environmentally sound, and locally controlled [1]. A well-known example of the appropriate technology is Q Drum, a plastic water tank which can be easily moved with its rotation [2].

B. Assisting local designers

So far, the designers of these problem solving products needed to understand the culture and the situation in the area they are targeting thoroughly since they were usually the outsiders of the community. To achieve this purpose, for instance at the d.school in Stanford university, various methods were invented for the designers to discover the local needs efficiently. To identify the essential needs in the developing countries, the framework of the *participatory development* was contrived in the field of the international development, which proceeds the project through interaction with inhabitants [3]. The *inclusive design* concept is also similar, to include the people, such as physically challenged people, into the design process to identify the socially important needs [4]. On the other hand, in terms of sustainability, it is also important to establish the environment for the local designers to design the product by themselves. For example, the tutorial to make some appropriate technologies is provided online for the local people to recreate the products by themselves, which is called Open Source Appropriate Technology (OSAT) [5]. Moreover, in the *Fab lab*, the worldwide open workshop which provides the digital fabrication tools to the people in the community, local designers have produced some useful products. For example, the Fab-Fi, the low cost Wi-Fi antenna for building the Internet mesh network, was developed in the Fab lab Jalalabad in Afghanistan [6]. Despite the increase of those platforms, the authors believe there still are plenty ways to accelerate the creation of the problem solving products in the developing countries. In this paper, as shown in Figure 1, we propose the design method and the support system for local designers to create the problem solving products by themselves.

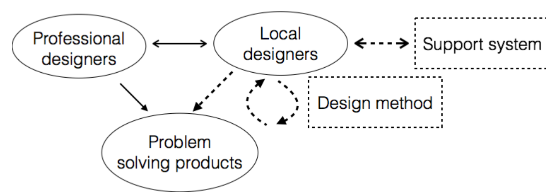


Figure 1. Design method and support system for assisting local designers

First of all, in Section 3, we discuss the existing difficulties that local designers face and clarifies the requirements for creating the design method. Considering the result, we propose the design method based on our past projects in Section 4. Moreover, we present the support system for assisting the design using our method in Section 5.

II. RELATED WORK

A. Appropriate technology

A lot of products of the appropriate technology have been invented and currently in use [2]. For example, a solar cooker is a device to generate heat using solar power as an energy source. People can make the reflector of the solar cooker with an unused parabolic antenna, which is an example of using materials in a different way. Other example is a hydraulic ram, a kind of water pump which can lift the water to a higher point without any external energy source. The hydraulic ram is interesting because it is based on a simple principle of fluid dynamics, which is the water hammer. These products

 ISBN: 978-1-61208-325-4

were designed by the experts who have the knowledge of each domain, therefore we aim to involve more local people in the design of such products. In this perspective, some successful products designed by ordinary people in South Asian countries have got attention. These products are called as frugal innovation or *Jugaad*, which means a creative problem solving with improvised arrangements [7]. For example, a hand-made vehicle was invented by local inhabitants using the agricultural water pump. Although there are such cases of the design by local people, there seems to be little researches on the design methodology and the support system for it.

B. Prototyping for creating ideas

One of the key issues of creating the problem solving products is to identify the needs of local people correctly. Some studies clarified the effectiveness of the prototyping for discovering the personal needs. Lim et al. pointed out their unique prototyping method called Discovery-Driven Prototyping (DDP) promoted discovering underlying issues of individuality [8]. Kelly also mentioned the importance of the prototyping in the problem solving [9]. So far, the prototyping of hardware has taken a long time; however, the recent developments of the fabrication tools drastically changed the environment around amateur creators. A 3D printer enabled the production of the plastic parts with less cost and time. With a laser cutter, the creators can cut the wooden or acrylic board in precise dimensions. Moreover, the micro controller board such as Arduino enabled users to prototype the electronics much easily [10]. Using such micro controller board, users can try various combinations of sensors and actuators to explore the design solution. We implement the support system using those fabrication tools effectively. It is important to support both conceptualisation and manufacturing in the design process since our targets are local designers, who usually don't have much knowledge and experience in design and manufacturing.

C. Creativity support

Boden described the three forms of creativity [11]. The first one is the combination of familiar ideas in unfamiliar ways. The second and third are the exploration and the transformation of conceptual space. Gero divided the design into the three types: routine design, innovative design and creative design [12]. Their arguments seem similar if we define the creative design as the transformation of conceptual space. Finke et al. discussed the creative activity based on the psychological experiment [13]. They proposed *geneplore model*, which demonstrates that the creative process is a cyclic process, which consists of generation of preinventive structure and preinventive exploration and interpretation. Although such theoretical researches and the laboratory experiments exist, there are a small number of practical studies on how to assist local designers to make the problem solving products in developing countries. Based on those preceding studies, we propose the design method, which promotes the exploration of the design space through the continuous prototyping.

III. Requirements Analysis

A. Identification of needs

It may seem easy for a local designer to identify the needs in a community, and no special design method is necessary to identify the needs. However, from the experience in the past project, we propose a design method is required not only for external designers, but also for local designers. In 2010, the authors engaged in the project with a Non-Governmental Organization (NGO) to develop a movie material of the science experiment for enhancing the educational quality at a farm area in Bangladesh. The movie was shot in the high school known for its high educational level. Contrary to expectations, the created movie was not good enough to be published in terms of the quality of the content, mainly because of the lack of the safety instruction by the teacher. The interesting point was that, prior to making the pilot movie, no one in the local NGO was suspicious of the quality. In this sense, making the prototype enabled us to realise the hidden assumption in the community, which was the less awareness of the safety education. However, the more interesting point was that one of the members in the NGO finally proposed to reuse the movie for the teachers not for the children, in order to enlighten the teachers about the importance of safety instruction. We point out it is a good example of the transformation of the product use to a different way. Every single person, no matter where he or she is from, has the personal bias based on the experience or cultural background, which unconsciously regulates his or her thoughts. Liberating designers from those unconscious constraints makes it possible to explore the different solutions unexpected at the initial phase in the design process.

B. Considering limitation of materials

In general, manufacturing resources, such as materials and tools, are limited in the developing countries. These constraints sometimes make designers give up using specific material. However, the limitation of materials sometimes promotes designers to come up with an innovative way of solving problems. For example, in our project in Ghana, we have been forced to select a wooden board rather than metal for making a magnetic rotor because of the availability. As a result, the choice enabled us to make the parts easier and with less cost. In another instance, the Jaipur Foot, a prosthesis foot developed in India, uses locally available rubber which satisfied the various requirements at the same time, such as the resemblance to the appearance of normal human foot, durability, waterproofness, and affordability [14]. The product diffused from India into parts of the world, which is opposite of the conventional technology dissemination. Recently, the movement of these inventions is referred as the reverse innovation [15]. Regarding this, we need to focus on the positive aspect of the limitation of materials.

C. Manufacturing knowledge

The open source software and hardware made it easier to develop the products with less cost, time and human resources. Besides, the online tutorial, such as *Instructables*, also enabled the sharing of knowledge and skills of the manufacturing. However, so far the designer of the appropriate technology was the expert who has the fundamental understanding of the technology. Since we aim to assist the local designers including amateurs, not only providing knowledge but also supporting the utilization of the knowledge is required.

IV. DESIGN METHOD

A. Definition of design space

Conventional design studies provided several ways for describing design space to discuss design process. For example, Maher et al. presented the co-evolution model, which describes the design exploration as the interaction of problem space and solution space [16]. As shown in Figure 2, we split the design space into four layers which are problem space, function space, structure space, and tool and material space. The problem space consists of the needs of the product users. The function space involves the required function of the products. The structure space is the collection of the entities which materialises the given functions. The tool and material space represents the whole available materials and the facility of manufacturing including machine tools. This separation concept is similar to the design model proposed by Suh [17]. In particular, we focus on the feedbacks from the tool and material space to the other spaces, which we believe is the unique feature of production in the developing countries. Due to the limited resources, the exploration of the suitable structure is regulated by the locally available materials. We use this design space model to discuss the transformation of each space during the design process.

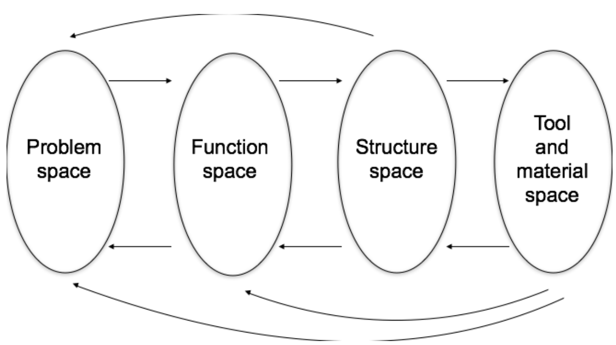


Figure 2. Definition of the design space

B. Dynamic transformation of design space

The main focus of the proposing design method is to transform the design space dynamically by the iteration of the prototyping. Here, we show the transformation of the design space taking the project in Ghana as an example.

In Figure 3, the solid arrows represent the initial flow of the designer's thoughts. In our project in Ghana, we initially targeted to solve the e-waste problem that was associated with the health hazard among the laborers who are working in the waste dump to extract metals such as gold and copper by burning the wastes. To avoid such problem, we defined the required function as the eddy current separation, which can separate the metals from plastics without burning the wastes. Subsequently, we picked up the sub functions of the eddy current separation as follows:

- Magnetic rotor to generate the electromagnetic field by the rotation of the magnets
- Power source for driving the magnetic rotor

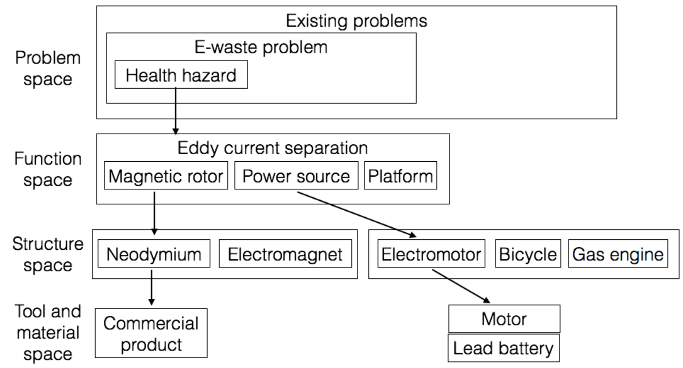


Figure 3. Design flow at initial state

- Platform to hold the magnetic rotor, a drive shaft, and bearings

Then, we started designing the magnetic rotor. At the beginning we expected to use the neodymium magnets, but later we noticed there were no neodymium magnets available in the local shops. After discussions, we decided to dismantle hard disk drives to extract the neodymium magnets. This transition of the design can be regarded as the extension of the tool and material space because we were able to discover the availability of the particular material through the design process. Such transformation is shown in Figure 4.

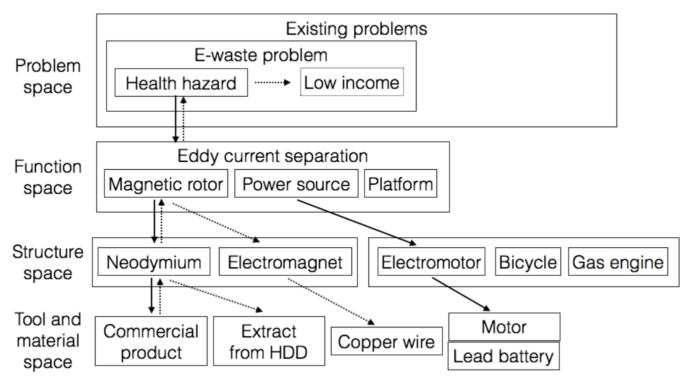


Figure 4. Transformation of structure space and tool-and-material space

In Figure 4, the dotted arrows show the flow of the induced thoughts through the prototyping process. In similar ways, each design space reflects on other design spaces, which triggers the transformation of the design space. Interestingly, as we mentioned the uniqueness of the tool and material space, the tool and material space can affect the problem space directly.

For example, as indicated in Figure 5, if we come up with the idea of using the electric motor and the lead battery to drive the magnetic rotor, it signifies that we are aware of the local availability of the lead battery. Then, we can start the discussion on the possible product made with the lead battery, which may lead to the use of the product to supply the electricity in the unelectrified area.

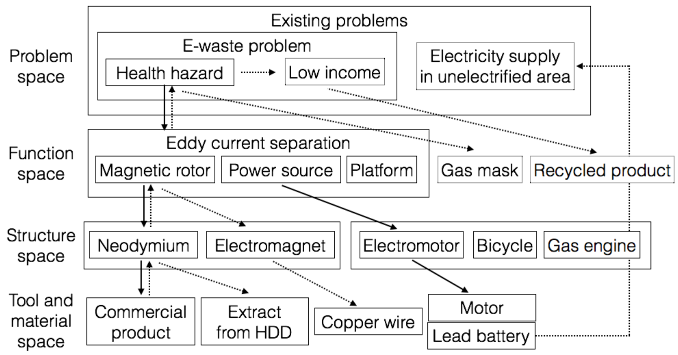


Figure 5. Transformation of function and problem space

C. Rapid prototyping spiral

Our focus is how we can accelerate the transformation of the design space. We propose the design method using the quick iteration of design, manufacturing the prototype and testing, which we call *rapid prototyping spiral*. The significance of prototyping is described in the design field from various perspectives. For example, Thoring et al. mentioned that the iteration of prototyping enables designers to find the next problem, while explaining the process of the design thinking [18]. In Figure 6, our design method is shown compared with the design thinking process.

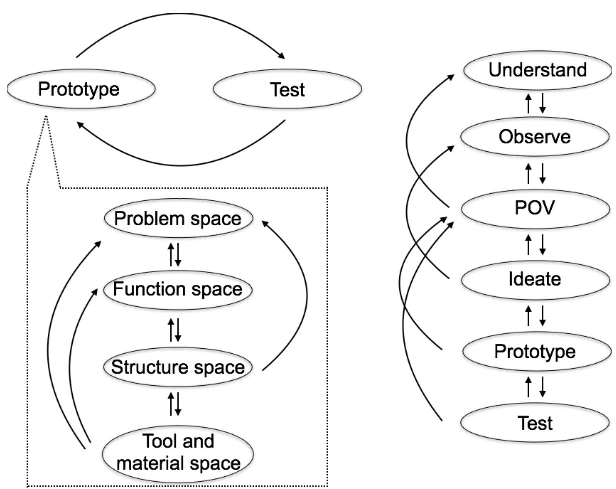


Figure 6. Prototyping spiral compared to the design thinking (used the figure from HPI D-School as reference [18])

In Figure 6, the left part shows our rapid prototyping spiral, and the right side represents the design thinking process. As shown in the figure, we propose to start the design process from prototyping. The reason we changed the order in the design process is that we focus on the following three aspects, which should be considered when assisting local designers in the developing countries.

- Limitation of materials
- Intrinsic motivation
- Production of practical products

TABLE I. EXPLORATION OF FUNDAMENTAL PROBLEMS

Design phase	Problem space	Prototype
Initial phase	Low quality of science education Lack of resources	Science movie shot in the high-level school
Second phase	Lack of carefulness for safety instruction	Material for teachers to instruct safety

First, due to the limitation of materials, local designers may not be able to manufacture the product with the materials required in the ideal design. In our method, by starting the design process from prototyping, the designers are promoted to focus on the material space mentioned in the previous section, and it leads to the exploration of the unexpected use of materials. Second, we expect that the continual process of the materialisation motivates the local designers, and such motivation enhances the creativity in the design process as indicated in the conventional researches [19] [20]. Third, the problem solving in the developing countries requires not the conceptual but the practical outputs as materialised things. Although the primary role of the prototype in our design method is an explorer of the design space, it is also aimed to be used as the real solution at the same time.

V. SUPPORT SYSTEM

We propose the support system to assist the design using our method. The support system consists of following four components.

A. Exploration of fundamental problems

What makes our design method differ from the conventional product development is the transformation of the initial problem space during the design process. The purpose of transforming the problem space is to find out the fundamental problem, in which the initially addressed problem is rooted. We illustrate this feature by taking our project in Bangladesh as an example. As shown in Table I, we assumed the initial goal as creating the science movie to compensate the experimental environment at the farm area, which ended up revealing the relative incautiousness of the safety instructions even among the teachers at the school in an urban district. To accelerate such discovery of the fundamental issues, the iteration cycle of the prototyping will be effective.

B. Knowledge base of available materials and tools

The advantage of making knowledge base of the locally available materials is that it enables users to try the alternative solutions when the expected material is not available. The knowledge base is also helpful for reducing manufacturing time. For example in our project in Ghana, we faced difficulties regarding materials. When we looked for the metal shafts and pulleys, it took so much time to get the suitable ones because the local shop did not always stock fixed kinds of products. In this case, because the shops were concentrated in a specific commercial area, it was better to create the inventory for the whole area in advance. With the knowledge base, the designers can run the prototyping process more quickly. Furthermore, making the knowledge base of the available tools is also worthwhile. Although the fabrication tools are limited in the developing countries in general, each situation is inhomogeneous definitely. In our project in Ghana, we were able to use a laser cutter and a large CNC (Computer Numerical Control) milling machine, which seemed unusual equipments compared to the ones in the surrounding areas. The authors define those irregular environments as *technological spike*. We expect to create the products which are adaptable to the locality, by making the knowledge base of the tools including these technological spikes.

C. Knowledge base of manufacturing

The manufacturing knowledge base is required to support the innovative design [21]. Existing knowledge base to support the ordinary creators, such as Instructables, includes the list of the required materials and tools, and the description of the each process of manufacturing. Additionally, we propose to add the feature that the user can append additional knowledge on the different layer of the description. The additional knowledge is the knowledge discovered during the design process, for example, the description of a material which is different from its ordinary usage. Although the general documentation is expected to be simplified and include the least information required, we assert that these redundant information is valuable.

D. Archiving format of prototyping

We suggest the archiving format to describe the design process as shown in a following figure.

Prototype	Version 1	Version 2	Version 3	...
Problem space				...
Function space	Unstable rotation	Magnets fly apart		...
Structure space	Eccentricity of the rotor Asymmetrically-distributed casings	Covered circumference of the rotor with plastic sheets		...
Tool and material space	Less cutting accuracy →	Use large CNC milling machine	Use an oil can to cover the rotor	...

Figure 7. Archiving format of the prototyping

With the format shown in Figure 7, we are able to describe how each prototype affected the four design spaces. The reason why we relate the transformation of design space to each prototype is because it enables us to track such transformation in a chronological order. In the words of software development, this archiving format corresponds to the revision control.

VI. Ongoing Experiment

In this section, we explain how our method is applicable for assisting the design and manufacturing in the developing countries. We introduce the ongoing experiment that we are currently preparing with the help of the Takoradi Technical Institute (T.T.I.). T.T.I. is a technical high school in Ghana, which has a Fab lab inside. In the Fab lab, students can use digital fabrication machines, such as a laser cutter or a CNC milling machine. First, we explain how the designers can identify the fundamental problems using our design method.

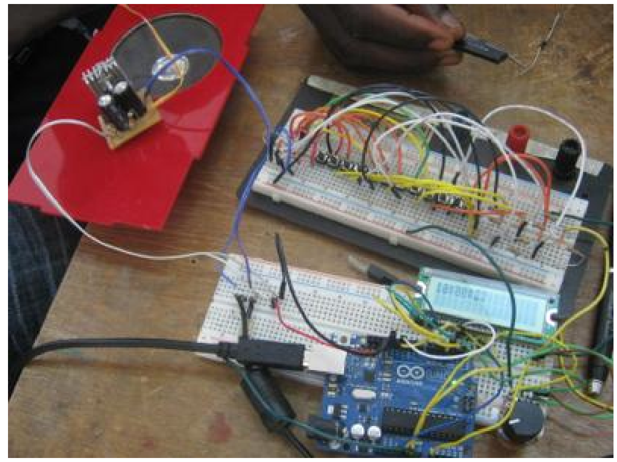


Figure 8. Sound alarm prototype

Figure 8 shows the prototype of a sound alarm made with our design method, which announces the beginning and end of the class to the teachers and students. It was based on the need of the principal, who told the authors the only one function to implement, which is emitting a sound triggered by a timer. The authors made the first prototype using Arduino, which plays a song at a scheduled time with electronic sound generated by a micro controller. The initial reaction of the teachers in T.T.I. to the prototype was different from expected, which was surprising to the authors. They demanded the authors to change the electronic sound to much louder and noisier siren. This feedback is noticeable because it changed the requirements specification of the prototype. Even though this observation is a slightly trivial example, by continuing to give feedback to the initial design space, we believe that the iteration process contributes to clarify the hidden assumption and fundamental problems.

Secondly, we introduce another example to discuss the transformation of design space. In this experiment, we set the goal as developing a 3D printer to produce inexpensive plastic parts. The expected transformation of design space is shown in Figure 9. In Figure 9, as well as the example shown in Section 4, the design proceeds from problem space to next spaces. A typical concept of a 3D printer using fused deposition method (FDM) is assembling an extruder of a plastic filament and a Computer Numerical Control (CNC) machine. Recently, the RepRap project allowed ordinary people to create a 3D printer by sharing all the information including 3D data of the parts and schematics of the circuit board [22]. Using such resources, in addition, we prepare the same mechanical unit for x, y, and z axes for building a CNC component. We take this module-based approach so that we can easily disassemble and reassemble it to build different machines. By creating different types of machines, we promote designers to discover unexpected functions of the machine. Also, in this example,

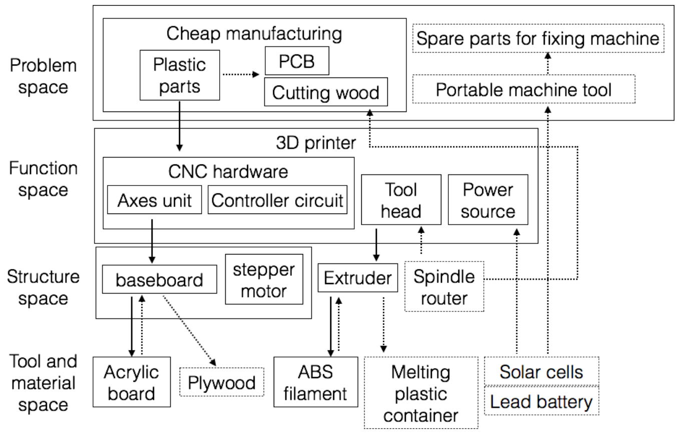


Figure 9. Expected transformation of design space

the tool and material space plays an important role. We try to use several materials, such as an acrylic board, plywood and even scrap wood, to find out the best choice in terms of cost and availability. The material space has an effect not only on the structure space, but also on the problem space. If a designer finds a lead battery, he or she may come up with the idea of powering the 3D printer by the battery. The customization makes the 3D printer portable, which can be regarded as the transformation of structure space. The portable 3D printer enables designers to take it to home and discover the possible use of it, such as creating the spare parts for fixing machines. This is an exploration of problem space, which is especially focused on the rapid prototyping spiral. After the whole design process, we analyse how the design space was transformed. When recording the design process, we utilize the web-based support system for archiving prototypes, as discussed in Section 5.

VII. Conclusion and Future Work

We proposed the rapid prototyping spiral, a design method for assisting the local designers in the developing countries to create the problem solving products. In our design method, the four kinds of design space are transformed during the design process, which enables us to discover the underlying issues in the users and the community, or come up with the unexpected usage of the materials. We also discussed the features of our support system to assist the design using the method we are proposing. As an example of applying our method to the design project, we introduced the experiment which is currently going on in Ghana. Based on the example of creating a sound alarm prototype, we observed a change of the requirement of a prototype, which induces the transformation of the design space. In addition, we introduced the design project to create a 3D printer, to prove the iteration of prototyping process contributes to identify more fundamental problems. We believe that the rapid prototyping spiral can also be applied to design and manufacture in the developed countries. The core of the method is supporting designers who have limited resources, such as materials, tools, and manufacturing knowledge. With the increase of the so-called maker communities and availability of digital fabrication tools, more individuals in the developed countries will be involved in design projects in the near future. We expect that our method also assists individual creators in the developed countries to identify the socially important issues, which leads to user-oriented innovation.

References

[1] I. Smillie, Mastering the Machine Revisited: Poverty, Aid and Technology. Practical Action, December 2000.

[2] C. E. Smith, Design for the Other 90%. Cooper-Hewitt Museum, 2007.

[3] V. Desai and R. B. Potter, The Companion to Development Studies. Hodder Education, 2008.

[4] S. Keates, P. J. Clarkson, L.-A. Harrison, and P. Robinson, “Towards a practical inclusive design approach,” in Proceedings on the 2000 Conference on Universal Usability, ser. CUU ’00. ACM, 2000, pp. 45–52.

[5] J. M. Pearce, “The case for open source appropriate technology,” Environment, Development and Sustainability, vol. 14, no. 3, June 2012, pp. 425–431.

[6] N. Gershenfeld, FAB: The Coming Revolution on Your Desktop - From Personal Computers to Personal Fabrication. New York: Basic Books, 2005.

[7] N. Radjou, J. Prabhu, and S. Ahuja, Jugaad Innovation: Think Frugal, Be Flexible, Generate Breakthrough Growth. Jossey-Bass, April 2012.

[8] Y. kyung Lim, D. Kim, J. Jo, and J. bum Woo, “Discovery-driven prototyping for user-driven creativity,” Pervasive Computing, IEEE, vol. 12, no. 3, August 2013, pp. 74–80.

[9] T. Kelley, “Prototyping is the shorthand of innovation,” Design Management Journal (Former Series), 2001, pp. 35–42.

[10] J. Sarik and I. Kymissis, “Lab kits using the arduino prototyping platform,” in Frontiers in Education Conference (FIE), 2010 IEEE. IEEE, 2010, pp. T3C–1.

[11] M. Boden, The Creative Mind: Myths and Mechanisms. New York: Basic Books, 1991.

[12] J. S. Gero, “Computational models of creative design processes,” in Artificial Intelligence and Creativity, ser. Studies in Cognitive Systems, T. Dartnall, Ed. Dordrecht: Kluwer Academic Publishers, 1994, vol. 17, pp. 269–281.

[13] R. A. Finke, T. B. Ward, and S. M. Smith, Creative Cognition. MIT Press, 1992.

[14] D. H. S. Katherine Ott and S. Mihm, Eds., Artificial Parts, Practical Lives: Modern Histories of Prosthetics. NYU Press, April 2002.

[15] V. Govindarajan and R. Ramamurti, “Reverse innovation, emerging markets, and global strategy,” Global Strategy Journal, vol. 1, no. 3-4, 2011, pp. 191–205.

[16] M. L. Maher, J. Poon, and S. Boulanger, “Formalising design exploration as co-evolution,” in Advances in formal design methods for CAD. Springer, 1996, pp. 3–30.

[17] N. P. Suh, The Principle of Design. Oxford University Press, 1990.

[18] K. Thoring and R. M. Müller, “Understanding the creative mechanisms of design thinking: an evolutionary approach,” in Procedings of the Second Conference on Creativity and Innovation in Design. ACM, 2011, pp. 137–147.

[19] E. A. Edmonds and L. Candy, “Knowledge support systems for conceptual design: the amplification of creativity,” in HCI (2), 1993, pp. 350–355.

[20] T. M. Amabile, Creativity and Innovation in Organizations. Harvard Business School, 1996.

[21] J. C. Thomas, A. Lee, and C. Danis, “Enhancing creative design via software tools,” Communications of the ACM, vol. 45, no. 10, 2002, pp. 112–115.

[22] R. Jones, P. Haufe, E. Sells, P. Iravani, V. Olliver, C. Palmer, and A. Bowyer, “Reprap - the replicating rapid prototyper,” Robotica, vol. 29, no. 1, 2011, pp. 177–191.

Characteristics, Attributes, Metrics and Usability Recommendations: A Systematic Mapping

Franciela Nissola
Graduate Program in Applied Computer
UNIVALI – Universidade do Vale do Itajaí
Itajaí, Brazil
nissolainf@gmail.com

Fabiane Barreto Vavassori Benitti
Graduate Program in Applied Computer
UNIVALI – Universidade do Vale do Itajaí
Itajaí, Brazil
fabiane.benitti@univali.br

***Abstract*— The research scenario concerned about software usability addresses a wide range of characteristics, attributes and evaluation models. As result, we can observe the evaluators difficult in order to select the characteristics that best apply to the product that is being evaluated. By this way, this paper presents a systematic mapping in order to identify the main characteristics of usability evaluation related to web, desktop and mobile devices environments. As results, we selected 31 papers in order to perform data extraction and way possible to produce a list of 28 evaluation characteristics.**

Keywords-Usability Evaluation; Systematic Mapping.

I. INTRODUCTION

Usability is a characteristic of a certain product related to the ease of use, speed of learning, control and management of error, solving the tasks that proposes to carry out with efficiency, effectiveness and mainly offering a high degree of satisfaction for its users [1][2]. By relying on the type of user and the experience of this, as well as the tasks performed and the execution environment of the system, usability is a system property that is not constant [3]. Usability is a quality of use, measured for a given context in which software is operated. Thus, the same software can provide a good usability for an experienced user, but bad usability for a beginner, or vice versa, or it can also be easy to operate if the system is only used sporadically, but difficult when used frequently [4].

Despite all these concepts for usability, yet there is no consensus among the researches for a model that clearly describes all its features. This knowledge gap is one reason why the majority of usability studies are either simplified or specific (studies that are performed in certain contexts) [5]. Along with the various concepts related to usability and addressed in the researches, it is clear that usability is treated as a determinant of success or failure of any software product, because it allows to generate greater efficiency by the users performing their tasks, helps in reducing training costs and allows that the integration of software system in the work environment of the users can be made in a more positive and less traumatic way [6].

Another relevant aspect of usability is perceived by the vast amount of research found in this area. It has an extensive set of features, attributes and models of usability evaluation available [7][8][9]. Each of these valuation models presents its specific set of characteristics, and the same feature may have been addressed in several models or be disregarded. Another situation found in these publications are characteristics or attributes of usability presented with different nomenclatures, but with the same concept and similar concepts between models.

Note that the evaluation models are often refined and developed to a level where specific execution environments are considered for the evaluation of the usability [10]. However, even with this refinement of the models and attributes per execution environment, yet are different attributes of usability evaluation addressed to the same specific execution environment. The use of specific attributes of usability according to the execution environment makes the evaluation of the usability more consistent, provides a greater investment guarantee and also the professionals involved in the production of software for this environment can use these attributes as drivers for creating software with better quality [11].

All these different ways of evaluation emerged as a contribution to the evolution of usability evaluation, providing to the software development organizations a varied set of criteria that can be applied to specific project types [12]. Although it became evident through research that the introduction of usability practices in software development processes are beneficial, many companies oriented to development face difficulties to actually apply these practices, because it is a complex issue, which has evaluation methods underutilized and difficult to understand for development teams. This difficulty is manly in the definition of the model and attributes that should be applied in the evaluation process, due to the large amount of existing research [12][13].

In the research, the models have some limitations when compared with each other and still there is relatively little information to support the selection of a set of attributes to evaluate the usability by the evaluators. Moreover, it is not always clear how the usability attributes presented in the various models are more or less advantageous than others [8]. Based on this scenario, this article presents a systematic mapping on the characteristics and attributes of usability evaluation proposed in primary studies between the years 2005 and 2012 for the execution environments Web, Desktop and Mobile Devices, consisting an important step to meet, broadly, the dismembering of the usability concept, considered by many researches under different environments, being an important step in the reach for a model that integrates all aspects identified.

The terms: characteristic, attribute, and recommendation metric, cited in this article, are used to classify the various aspects, quality and information related to usability, obtained through this systematic mapping. It was chosen to use a hierarchical structure for the presentation of issues related to usability found in the articles evaluated, considering that other models in the literature also use hierarchical structure to present information related to usability (Model QUIM, [8]).

It´s assumed in this systematic mapping of the literature that the features are on top of the hierarchical structure mentioned, representing a more comprehensive view of the quality related to usability. The attributes are below the features and represent the information in a more granular form. In this systematic mapping a feature may or may not be related to one or more attributes and the same attribute may be related to more than one feature. Finally, on the basis of the hierarchical structure, are the metrics that represent the information used for the evaluation of the attribute more objectively, and the recommendations, which aims to guide the development/evaluation of interfaces through good practices, for example, layout and menu structure, pattern of presentation for links, font color for text, etc.

This article aims at presenting the results of a systematic survey of the literature involving characteristics, attributes and usability found recommendations, aiming to know, in an impartial manner, the state of the art concerning these aspects of usability, in relation to different platforms. Therefore, the protocol for a systematic mapping of the literature used in this paper is presented in Section II, along with the implementation procedures and discussion of the results. Finally, the last section summarizes the findings and presents some possible future work.

II. SYSTEMATIC MAPPING OF THE LITERATURE

The main reason to perform a systematic mapping of the literature is to increase the quality of material used about the subject of interest, increasing the success of the investigation, avoiding unnecessary duplication of effort and errors.

A systematic mapping follows a defined sequence of methodological steps, which in turn provide a high scientific value to the results. In execution of a systematic mapping, there is the establishment of criteria, search strategies and a description of all the elements previously required to hold the methodological conduction of the activities [14]. Thus, the following sections detail each step performed.

A. *Planning*

In order to map the attributes of usability evaluation in execution environments Web, Desktop and Mobile Devices, the following question was established: Which attributes are related to the usability of software products when considering different execution environments (among those contemplated in this research: web, desktop and mobile devices)?

To answer the research question there were four research bases; quote: (i) ACM Digital Library, (ii) IEEExplore Digital Library, (iii) ScienceDirect, and (iv) SpringerLink. At each base, the following search string was executed: (usability) AND (attribute OR factor OR criteria OR metric OR measurement OR method OR evaluation OR requirement) AND (web OR internet OR www OR desktop OR 'mobile device').

The language adopted in the survey was English, being the predominant language in the study area. Regarding the reliability of the content on those studies, it is considered that the works have already gone through preliminary review and assessment that allowed their inclusion in searchable databases. The process will be through search of primary studies published in proceedings, journals and periodicals available in international databases. After obtaining the references of the studies obtained in this four bases, some selection and exclusion criteria that would help to limit the items that could help to answer the research question were adopted (see Table I).

TABLE I. SELECTION CRITERIA AND EXCLUSION

Selection Criteria
SC1. Studies available on some web source (among the sources cited in this article);
SC2. Studies that meet at some level, the research question. Studies that show usability attributes, applied to a software product, related to at least one of the execution environments covered by the survey;
SC3. Studies that show evidence of implementation of the proposed attributes;
SC4. Studies that show the conceptualization of the proposed attributes (the concept can be in the primary study itself or pointed location/directed by the authors);
SC5. Complete studies published in proceedings and journals since 2005;
Exclusion Criteria
EC1. Studies that do not show attributes of usability in at least one of the execution environments covered by the survey or don´t make clear the execution environment;
EC2. Studies that do not conceptualize the attributes used;
EC3. Studies that show a specific application domain, such as e-government, e-commerce, or focus on a specific user profile, for example, elderly, impaired vision, etc.;
EC4. Summaries of studies will not be selected;
EC5. Studies that don´t show the evidence of application;
CE6. When a study has multiple publications in journals or proceedings, the most complete version of the study will be used;

B. *Conduction and Extraction*

The conduction of the systematic mapping was done from September 2012 until March 2013, consisting of the execution of the search protocol. The conduction process observed the steps shown in Figure 1, from 345 potential studies, 65 studies were selected. The conduction of the studies was done in three steps:

- Selection and cataloging preliminary studies collected: a preliminary selection of publications made from the application of the search string in the search sources selected.

- Selection of relevant documents: a preliminary selection using the search string does not guarantee that all the material that was collected is used in the research context. Thus, after the identification of publications obtained through the search engines, the studies were analyzed according to the criteria established for inclusion and exclusion.
- Extraction of information of relevant documents: after setting the final list of relevant documents, the necessary information related to the research objective is extracted.

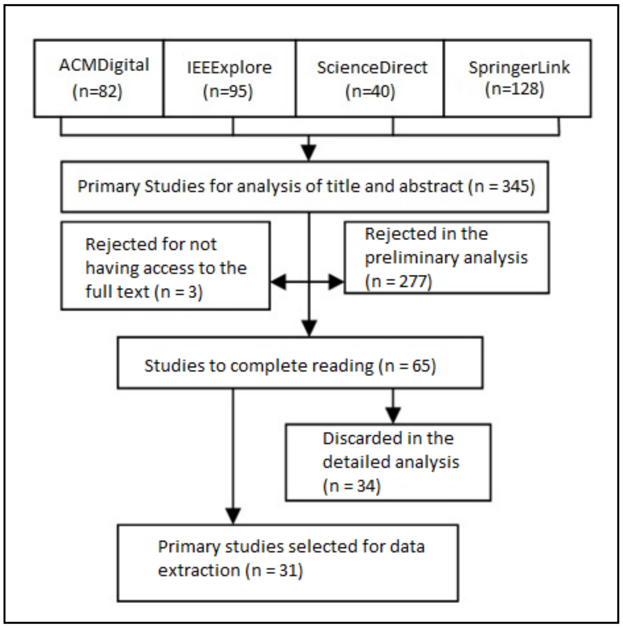


Figure 1. Process of selecting primary studies.

Among the 65 studies selected for full reading, we have the following scenario of events in each of the evaluated databases: IEEExplore Digital Library: 45,16%, ACM Digital Library: 22,58%, ScienceDirect: 19,35%, and SpringerLink: 12,91%.

From the set of primary studies selected to perform data extraction, the following information was extracted for analysis:

- Title: identifies the selected study.
- Year: provides a temporal view of the attributes of usability evaluation.
- Database: identifies the origin of the selected study.
- Execution Environment: identifies the execution environment for the attributes shown in the study evaluated. In this topic, the studies were classified as: (I) web, (II) desktop and (III) mobile device.
- Reference type: identifies the sources used to support the model of usability evaluation presented in the study.
- Structure addressed in the model: identifies the structure presented by the study evaluated, related to the presentation of the attributes of usability.
- List of attributes: compilation of the attributes of usability evaluation cited in the studies analyzed.
- Metric or recommendations: identifies whether the study has evaluate metrics or recommendations to guide the evaluation of usability.
- Assessment technique: identifies the technique used to evaluate the model or usability attributes addressed in each study rated. It was used the classification of evaluation techniques proposed by [4], as shown in Figure 2.

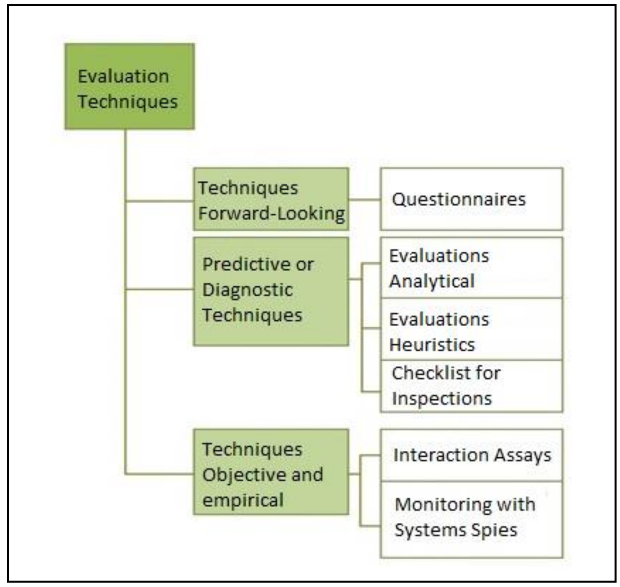


Figure 2. Usability evaluation techniques.

- Context: identifies the application context of the usability study. In this topic, the studies were classified as: (I)non-explicit,(II)academic/university and (III) business.
- Participants: identifies the number of respondents who rated the attributes or usability model presented.

In the process of selection of studies, the snowball technique [15] was applied. The application of this type of technique is to add in the research being conducted, studies referenced by the selected primary studies that may contribute to the research. Thus, when a primary study (among 31 selected) referenced another study that contains models, metrics and recommendations that could contribute to the research, this study referenced was also included. Appendix I is a listing of all studies, employing the technique of snowball nine studies were included (#35, #36, #37, #38, #39, #40, #41, #42, #43).

C. Results and Discussions

This section presents the main results obtained from 31 primary studies selected. Figure 3 shows the number of studies that were selected for data extraction in each year of the time interval considered in this systematic mapping (2005-2012). Over the years, it is observed that continuous studies have been published in this area and the research output remained, highlighting the importance of this area of research forward.

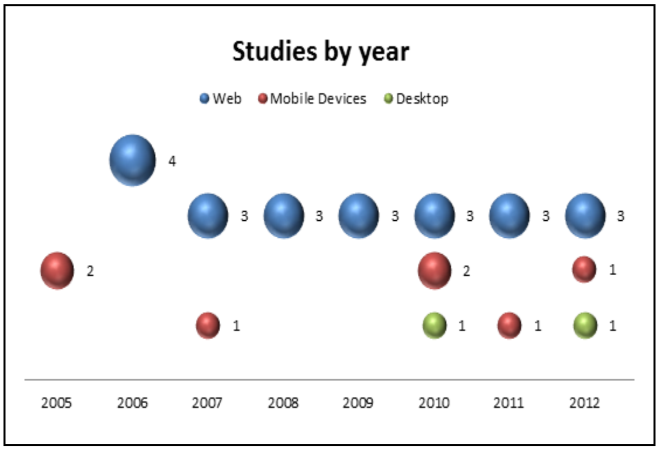


Figure 3. Number of evaluated annually.

In Figure 3, it can be seen that the occurrence of studies related to web runtime environment remained almost constant during the time period evaluated. This amount of studies related to web routine environment can be attributed, among other factors, to the grown of the Web, its popularization and its importance as a means of global communication.

In contrast, the execution environment desktop has the lowest incidence of studies to define a set of attributes to evaluate the usability. Among other factors, the rise of the Web previously mentioned, contributes to the unpopularity of the desktop environment and, therefore, to decrease research on this. In relation to mobile devices, it´s expected a growth of studies related to this execution environment, considering the growth of these environments (a report from Cisco predicts that mobile devices will exceed the number of people in 2013 and by 2017 there will be about 1.4 devices per capita) [16].

Regarding execution environments addressed in 31 studies evaluated, there´s the following scenario: 71% of the studies are related to the web environment, the mobile devices 23% and 6% related to the desktop environment.

As for the reference of the valuation models or list of attributes of usability evaluation presented in the studies evaluated, there´s the following scenario:

a) No reference model: 41,93% of the studies evaluated showed a list of attributes that were evaluated, using one of the techniques discussed in Figure 2, in a software product.

b) Adapted: 32.26% of the evaluated studies adapted a valuation model or list of attributes. In this group there is the use of ISO9241, ISO25000, and Nielsen heuristics[2].

c) Existence: 25.81% of the assessed studies used an existing model in the literature. In this group, it is important to highlight the use of ISO9241, heuristics and Nielsen Microsoft Usability Guidelines (MUG).

The scenario found as the reference of models or attributes of evaluation presented by this mapping, further enhances the already mentioned problem regarding the large amount of studies presented in which each author addresses their specific list of attributes and evaluates them not relating this list with models already existent, adapting or improving them.

It was also found that these 31 studies evaluated do not address a specific structure, for example, a framework model or categorization, to the listed attributes. The structure found in these studies, as the presentation attributes of usability evaluation is in the form of a simple list of attributes.

Regarding the application context of the studies evaluated, it´s observed that most of the selected studies did not indicate the context in which the attributes have been applied. Among the studies evaluated in this systematic mapping 74.19% do not make it clear which is the application context of the attributes addressed.

The context of business application was not used in any of the studies selected for this mapping. This scenario refers to the need to extend research regarding the evaluation of characteristics and attributes of usability in business contexts, facilitating the application, in a practical way, of the researches conducted.

During the evaluation of the articles selected for data extraction, it was checked for the presence of objective recommendations, to assist the evaluation of characteristics and attributes presented. However, the present scenario is not favorable regarding obtaining such advice. Of the 31 articles evaluated, 83.87% of them do not show metrics/objective recommendations for usability. Only 16.13% of the evaluated studies have some kind of metric/recommendation for usability. Table II presents the recommendations and metrics found in the studies evaluated in this systematic mapping. The objective recommendations found in the analyzed studies still have certain subjectivity (according to the evaluation of the authors of this article). It was observed that not all recommendations were related to some characteristic or attribute of usability, in this case it was chosen to categorize as “General Properties”.

TABLE II. RECOMMENDATIONS AND METRICS OF USABILITY

Characteristics/Attributes	Recommendations/Metrics
Prevention of errors	6 Recommendations [#3]
Integration of Communication	3 Recommendations [#22]
Attractiveness	3 Recommendations [#22]
From User Control	2 Recommendations [#22] 22 Recommendations [#43]
Security	4 Recommendations [#22]
Flexibility	2 Recommendations [#40] 2 Metrics [#5]
Effectiveness	6 Metrics [#5]
Efficiency	6 Metrics [#5]
Navigation/guidance	12 Metrics [#5]
Satisfaction	2 Metrics [#5]

 ISBN: 978-1-61208-325-4

Characteristics/Attributes	Recommendations/Metrics
Aesthetic design	52 Recommendations [#43]
Commands	40 Recommendations [#43]
Textual information	23 Recommendations [#43]
Message	34 Recommendations [#43]
Consistency interaction	33 Recommendations [#43]
Window	62 Recommendations [#43]
Interface/Layout	8 Recommendations [#43]
Font	2 Recommendations [#43]
Colors	4 Recommendations [#43]
Ícons	8 Recommendations [#43]
Animation and Transition	7 Recommendations [#43]
Graphic elements	7 Recommendations [#43]
Sound	8 Recommendations [#43]
User Experience	29 Recommendations [#43]
Components	65 Recommendations [#43]
General Properties	5 Recommendations[#22] 15 Metrics [#5] 6 Metrics [#3]

Among the 31 studies evaluated, 28 distinct characteristics of usability evaluation were found. Of these, one has the following scenario of events for execution environment:

a) *Web:* 28 characteristics;

b) *Desktop:* 5 characteristics;

c) *Mobile Devices:* 11 characteristics.

Appendix II has a full list of features and attributes per runtime environment, obtained through systematic mapping. In this appendix are presented the characteristics and attributes of usability evaluation which were obtained from the mapping studies conducted. One of the criteria for selection of studies was presenting the conceptualization of the attributes used to evaluate the usability. Based on this concept, we observed attributes with the same concept, but with different nomenclatures in different studies and group these attributes into a single information structure.

It was also observed that some studies had a more comprehensive view of the quality of information related to usability, for these cases, we used the characteristics structure, which, in turn, is a sort of aggregator of information. The characteristics of the structure in Appendix II were present in at least one study evaluated, or when a more comprehensive study presented information structure composed of more granular information (attributes), this structure was maintained.

In this table, using a mark with the letter "x", the possible relationships between features and attributes discussed by several authors in the studies evaluated are also presented. A nomenclature of the characteristics and attributes discussed in studies with some minor adjustments when the same characteristic or attribute was discussed in more than one study in which a different nomenclature was showed, but with the same concept presented, was used.

In Appendix II, along with the naming of features and attributes, there is also the presentation of information from references that cited such characteristics or attributes (references are identified with the hash (#) concatenated with the numbering of the bibliographic reference used) facilitating the reader find an indication of which study addressed what characteristic and attributes, and you can also check the amount of referrals every feature or attribute has. The reference for the primary study selected in which the feature or attribute was found is also mapped in this table.

There are several ways to evaluate the usability of an application and, in the context of this article, the 31 studies evaluated were classified in Figure 2. The result, illustrated in Figure 4, shows the number of studies related to the use of the techniques of usability evaluation in each execution environment evaluated. The application of the questionnaire assessment technique was the most used among the studies evaluated. This type of technique is based on questionnaires to evaluate the user satisfaction or dissatisfaction with the system and its operation. It is a technique quite relevant, considering that the user is the person who knows more the software, and because of that, nothing is more befitting than to seek their opinions to guide project reviews [4].

Trials of interaction were also widely used, the second among the most widely used techniques for the assessment of the selected studies. It has been in the execution environment web site for greater application of techniques for usability evaluation cited. A test of interaction consists of a trial use of the system which involved subjects representative of the target population to evaluate the technique. In tests of interaction, participants perform typical tasks involved in their activities with a version of the desired system [4].

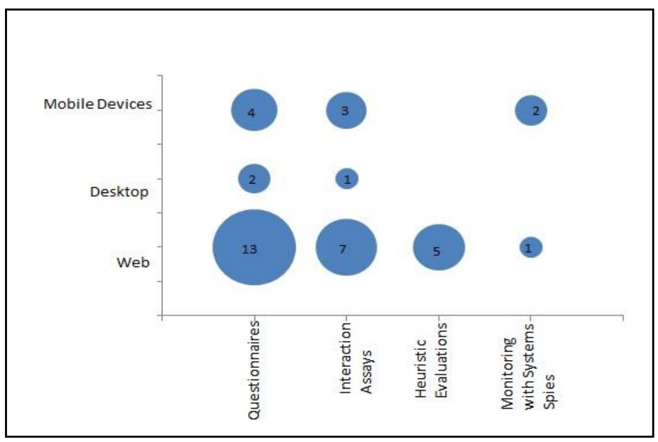


Figure 4. Evaluation Techniques for Runtime Enviroment.

Along with the information extraction of the usability evaluation techniques used in the selected studies, it was verified the number of people involved in these evaluations. It was found that such evaluations occurred either on a list of attributes or on a model (consisting of features and attributes) of usability evaluation and the number of people involved in these evaluations did not follow a pattern in relation to the quantity displayed. For the group of studies that evaluated a list of attributes, there are since evaluations with groups of four people to evaluations with groups consisting of 215 people. The same is true for the group of studies that evaluated a model of usability, working with groups with 7 people to groups with 179 people involved in the usability evaluation proposed.

In both cases (List of Attributes or Model), the evaluation technique most used was the prospective technique of evaluation by using questionnaires, as already mentioned. This technique is based on questionnaires to those involved in the evaluation and is facilitated with a view that those involved respond to a questionnaire on a specific web address and may be performing this evaluation activity anytime. However, it´s necessary to emphasize that satisfaction questionnaires have a low return rate of responses, which indicates the need to develop/use a small number of questions, as well as using succinct questions and a space for the user to insert free comments and suggestions [4].

III. CONCLUSION AND FUTURE WORK

The systematic mapping of the literature described in this article has covered features, attributes and metric/usability recommendations found in each of the execution environments evaluated in studies found between 2005 and 2012 in four databases of the area.

This systematic mapping of the literature has shown the large amount of features and attributes discussed by several authors in various studies related to usability, considering that it was verified a total of 28 features and 76 different attributes that can be used by those involved with a usability evaluation in one of the environments mentioned.

The great quantity of features and attributes of usability evaluation is also diverse when studies for the same execution environment are being verified. Among the studies reviewed, it wasn´t found a pattern or a set of characteristics and unique attributes of evaluation among the authors that check specific execution environments. In addition, different terms to refer to the same usability attribute are found, and the construction of an ontology is a perspective for future work.

Few studies have metrics or objective recommendations to assist in the evaluation of usability. Among the studies evaluated (31 selected and 9 added by snowball), only about 12.5% have metrics or recommendations to assist, through good practices, the development of interfaces.

Thus, the choice of features and attributes that best apply to a particular evaluation, among this diverse group found, is not a trivial task, given that often the people involved with the evaluation do not have necessary expertise in the area of usability that can assist and facilitate the evaluation work and, therefore, the make the best choice.

Still, when the people involved with the evaluation of usability have a set of characteristics and attributes to evaluate, sometimes it is not clear what and how the interface should be evaluated or even how the interface should be constructed to meet certain characteristics and attributes.

Given this scenario of complexity in the choice of features and attributes of usability, companies often do not apply methods of usability evaluation in their software projects/products for not having specialized team to select the best form of evaluation, as well as the best attributes for evaluation, producing sometimes software with low usability.

The results of this systematic mapping clarify even more the need for a model of usability evaluation to guide those involved in the assessment, to provide objective guidelines for the evaluation of interfaces and indicate the execution environment for which those guidelines are valid. Another perceived need is related to the application of usability evaluation in business contexts, favoring the practical application of the researches.

It is necessary a model or tool that assists in the process of evaluation and selection of evaluation attributes, facilitating the application of usability evaluations of projects/software products.

Finally, it is important to analyze the threats inherent to the process adopted in this research, which mainly consists in the disposal of three studies for not having access to the full text. Another threat is about the mapping conduction, which was performed by a single investigator, but the planning stages and discussion involved two researches. To mitigate this threat, it was performed, after the conducting conclusion, an analysis of the disposed titles, and in specific cases, an abstract by a second researcher (as a way of review) was made. It´s also necessary to consider that the result presented is restricted to a specific search string in a set of four data sources limited to search for terms in English. The expansion of the terms, language and databases can reveal new features, attributes, metrics, and recommendations.

REFERENCES

[1] International Organization for Standardization. ISO 9241-11:1998, “Ergonomic requirements for office work with visual display terminals (VDTs), Part 11: Guidance on usability. Geneva, Switzerland: 1998.

[2] J. Nielsen, Usability Engineering, London, UK: Academic Press, 1993.

[3] B. Shackel and S. Richardson, Human factors for informatics usability-background and overview, 1rd ed., vol. 1. Cambridge University Press: New York, 1991, pp. 1–19.

[4] W. de Abreu Cybis, Usability Engineering: An ergonomic approach, LabUtil: Laboratorio de Utilizabilidade de Informatica, 2003.

[5] E. Mosqueira-Rey, D. Alonso-Ríos, and V. Moret-Bonillo, “Usability taxonomy and context-of-use taxonomy for usability analysis,” Proc. IEEE Symp. Conference on Systems, Man and Cybernetics (SMC'09), IEEE Press, Oct. 2009, pp. 812-817, doi: 10.1109/ICSMC.2009.5346929.

[6] A. M. S. Filho, “Human Perception in Human-Computer Interaction,” in Revista Espaço Acadêmico, vol. 25, 2003.

[7] S. Winter, S. Wagner, and F. Deissenboeck, Engineering Interactive Systems: A Comprehensive Model of Usability, vol. 4940. Springer-Verlag: Berlin, Dec. 2008, pp. 106–122.

[8] A. Seffah, M. Donyaee, R.B. Kline, and H. K. Padda, “Usability measurement and metrics: A consolidated model,” Kluwer Academic, vol. 14,Jun. 2006, pp. 159-178, doi: 10.1007/s11219-006-7600-8.

[9] O. Signore, “A comprehensive model for Web sites quality,” Proc. IEEE Symp. Web Site Evolution (WSE 2005), IEEE Press, Sep. 2005, pp. 30-36, doi: 10.1109/WSE.2005.1.

[10] A. Fernandez, E. Insfran, and S. Abrahão, “Usability evaluation methods for the web: A systematic mapping study,” Information and Software Technology, vol. 53, Aug. 2011, pp. 789-817, doi: 10.1016/j.infsof.2011.02.007.

[11] R. Fitzpatrick, “Additional Quality Factors for the World Wide Web,” Proc. and Cybernetics – Part A: Systems and Humans, 1999, pp. 29-386.

[12] C. Lallemand, “Toward a closer integration of usability in software development: a study of usability inputs in a model-driven engineering process,” ACM SIGCHI symposium on Engineering interactive computing systems (EICS '11), Jul. 2011, pp. 299-302, doi: 10.1145/1996461.1996541.

[13] J. O. Bak, K. Nguyen, P. Risgaard, and J. Stage, “Obstacles to usability evaluation in practice: a survey of software development organizations,” Proc. Nordic conference on Human-computer interaction: building bridges (NordiCHI '08), IEEE Press, Oct. 2008, pp. 23-32, doi: 10.1145/1463160.1463164.

[14] B. Kitcheman, Guidelines for performing Systematic Literature Reviews in Software Engineering. Durham, UK: University of Durham, 2007.

[15] S. Jalali and C. Wohlin, “Systematic literature studies: database searches vs. backward snowballing,” Proc. ACM-IEEE international symposium on Empirical software engineering and measurement (ESEM '12), IEEE Press, Sep. 2012, pp. 29-38, doi: 10.1145/2372251.2372257.

[16] Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017, 2013.

Appendix I – Primary Studies Included in the Mapping

[#1]G.Lindgaard, C.Dudek, D.Sen, L.Sumegi, P.Noonan," An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages"(2011) [#2]R.West,K. Lehman, "Automated summative usability studies: an empirical evaluation"(2006) [#3]A.T.Neto, T.J.Bittar, R.P.M.Fortes, and K.Felizardo, "Developing and evaluating web multimodal interfaces - a case study with usability principles"(2009) [#4]A.Seffah, M.Donyanee, R.B.Kline, H.K.Padda, "Usability measurement and metrics: A consolidated model" (2006) [#5]P.A.Davis, F.M.Shipman, "Learning usability assessment models for web sites" (2011) [#6]C.Ryan, A.Gonsalves, "The effect of context and application type on mobile usability: an empirical study"(2005) [#7]A.Fernandez, S.Abrahão, E.Insfran, "Towards the validation of a usability evaluation method for model-driven web development"(2010) [#8]A.P.Massey, V.Khatri, V.Vamesh, "From the Web to the Wireless Web: Technology Readiness and Usability"(2005) [#9]A.Anandhan, S.Dhandapani, H.Reza, K.Namasivayam, "Web Usability Testing ? CARE Methodology"(2006) [#10]T.Tirapat, T.Achalakul, "Usability Assessment for Hyperlink Methods"(2006) [#11]T.Conte, J.Massolar, E.Mendes, G.H.Travassos, "Usability Evaluation Based on Web Design Perspectives"(2007) [#12]A.P.Massey, V.Khatri, M.M.Montoya-Weiss, "Online Services, Customer Characteristics and Usability Requirements"(2008) [#13]H.Y.Kao, "Usability Testing of AMC Hospital's Website for Home Users: Case Study for On-Site Registration Design"(2007) [#14]A.Granie, I.Mitrovic, N.Marangunie, "Usability Evaluation of Web Portals"(2008) [#15]T.Conte, J.Massolar, E.Mendes, G.H.Travassos, "Web usability inspection technique based on design perspectives"(2009) [#16]U.K.Yusof, L.K.Khaw, H.Y.Chang, B.J.Neowa, "Balancing between usability and aesthetics of Web design"(2010) [#17]A.Al-Wabil, L.Al-Husian, R.Al-Murshad, A.Al-Nafjan, "Applying the Retrospective Think-Aloud Protocol in Usability Evaluations with Children: Seeing Through Children’s Eyes"(2010) [#18]W.F.W.Ahmad, S.Sulaiman, F.S.Johari, "Usability Management System (USEMATE):A Web-Based Automated System for Managing Usability Testing Systematically"(2010) [#19]M.Freeman, M. Bowden, "Usability as a Panacea"(2010) [#20]L.Punchoojit, T.Chintakovid, "Influence of Age Group Differences on Website Cultural Usability"(2012) [#21]M.A.Mayz, D.M.Curtino, A.de la Rosa, "Avoiding laboratories to collect usability data: two software applications"(2012) [#22]A.Oztekin, "A decision support system for usability evaluation of web-based information systems"(2011) [#23]X.Fang, C.W.Holsapple, "An empirical study of web site navigation structures' impacts on web site usability"(2007) [#24]A.Fernandez, S.Abrahão, E.Insfran, "Empirical validation of a usability inspection method for model-driven Web development"(2013) [#25]M.Allen, L.M.Currie, S.Bakken, V.L.Patel, J.J.Cimino, "Heuristic evaluation of paper-based Web pages: a simplified inspection usability methodology"(2006) [#26]C.K.Coursaris, K.Hassanein, M.M.Head, N.Bontis, "The impact of distractions on the usability and intention to use mobile devices for wireless data services"(2012) [#27]E.K.Delice, Z.Gungor, "The usability analysis with heuristic evaluation and analytic hierarchy process"(2009) [#28]A.Holzinger, M.Errath, "Mobile computer Web-application design in medicine: some research based guidelines"(2007) [#29]M.Vigo, A.Aizpurua, M.Arrue, J.Abascal, "Quantitative assessment of mobile web guidelines conformance"(2011) [#30]W.Oyomno, P.Jappinen, E.Kerttula, K.Heikkinen, "Usability study of ME2.0,"(2012) [#31]D,Schreiber, M.Hartmann, F.Flentge, M.Mühlhäuser, M.Görtz, Thomas Ziegert, "Web based evaluation of proactive user interfaces"(2008) [#35]J.Nielsen, "Usability Engineering"(1993) [#36]ISO/IEC FDIS 25010, "Systems and software engineering – Systems and Software Quality Requirements and Evaluation (SQUARE) – System and software quality models"(2010) [#37]MICROSOFT, "Windows Enviroment" [#38] T.V.Raman,"Design principles for multimodal interaction"(2003) [#39]L.M.Reeves, J.Lai, J.A.Larson, S.Oviatt, T.S.Balaji, S.Buisine, P.Collings, P.Cohen, B.Kraal, J.C.Martin, M.Mctear, T.Raman, K.M.Stanney, H.Su, Q.Y.Wang, "Guidelines for multimodal user interface design"(2004) [#40]A.Fernandez, E.Insfran, S.Abrahão,"Integrating a Usability Model into Model-Driven Web Development Processes"(2009) [#41]J.M.C.Bastien, D.L.Scapin, "Ergonomic Criteria for the Evaluation of Human-Computer Interfaces"(1993) [#42]ISO9241-11,"Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability" (1998) [#43]K.Keeker, Improving Web Site Usability and Appeal, 1997".

Appendix II – List of Characteristics and Usability Attributes Mapped

(green)	Desktop
(yellow)	Mobile Devices
(blue)	Web

ATTRIBUTES \ CHARACTERISTICS	Integration of Communication #22	From User Control #41 #19 #22 #23 #35 #3 #4 #7 #11 #14 #15 #19 #24 #27	Navigation / guidance #5 #22 #28 #41 #19	Recursos visuais	Efficiency #42 #2 #35 #3 #40 #4 #5 #6 #7 #9 #10 #11 #13 #14 #16 #17 #18 #19 #20 #21 #24 #26 #27 #29 #30 #31 #6 #17	Effectiveness #42 #2 #40 #4 #5 #7 #10 #11 #13 #14 #17 #18 #20 #21 #24 #26 #29 #30 #31	Satisfaction #42 #2 #40 #4 #5 #6 #7 #10 #13 #14 #16 #17 #18 #20 #21 #24 #26 #29 #30 #31	Visibility of System Status #35 #3 #4 #7 #11 #14 #15 #19 #24 #27	Flexibility #35 #3 #5 #7 #11 #14 #19 #24 #27	Consistency interaction #39 #3 #28 #41 #43	Multimodal #3	Mapping between system and real world #35 #3 #4 #7 #11 #14 #15 #19 #24 #27	Liberty #35 #3 #4 #7 #11 #14 #15 #19 #24 #27	Minimalist #35 #3 #4 #7 #11 #14 #15 #19 #24 #25 #27	Help #35 #3 #4 #7 #11 #14 #15 #19 #24 #27	Documentation #35 #3 #4 #7 #11 #14 #15 #19 #24 #27	Management Errors #41 #19	Ease of use #4 #6 #7 #8 #9 #12 #24 #27	Productivity #40 #4	Learnability #40 #4 #6 #16 #20 #35 #3 #36 #4 #7 #11 #14 #15 #41 #19 #24 #27	security #40 #4 #22 #35 #3 #4 #7 #11 #14 #15 #41 #19 #24 #25 #27	Trustworthiness #40 #4 #16 #22	Accessibility #40 #4 #36 #7	Universality #40 #4	Utility #40 #4	Workload (Memory) #41 #19 #25 #30	Adaptability #6 #41 #19 #8 #12 #30 #36 #4 #7	Operability #41 #19 #36 #4 #7
Feedback #41 #19 #40 #4 #39 #3 #8 #12			x		x	x					x							x						x	x			
Load minimal memory #40 #4					x		x													x			x	x	x			
Operability #40 #4					x		x															x	x		x			
Minimal Action #40 #4					x		x													x			x					
Navigability #40 #4					x	x																x	x	x				
Flexibility #40 #4 #41 #19						x	x																x	x	x		x	
Charge Time #40 #4					x														x					x	x			
Consistency #40 #4						x															x		x	x				
User guidance #40 #4							x													x			x	x				
Self-Description #40 #4																				x		x	x	x				
Simplicity #40 #4																				x			x	x				
Controllability #40 #4																						x	x	x	x			
Privacy #40 #4																						x		x	x			
Legibility #41 #19 #40 #4			x																				x	x				
Temporal behavior #40 #4					x														x									
Use of resources #40 #4					x														x						x			
Accuracy #40 #4						x															x				x			
Completeness #40 #4						x															x							
Enjoyability #40 #4							x																					
Familiarity #40 #4																				x		x						
Fault Tolerance #40 #4																					x	x			x			
Safety Features #40 #4																					x				x			
Assurance #40 #4																					x	x						
User Experience #43 #41 #19				x																							x	
Density #1 #41 #19				x																						x		
Customization #23 #8 #12				x																							x	
Attractiveness #36 #4 #22 #7 #40				x			x																	x				
Explicit user action #41 #19		x																										
Solicitation #41 #19			x																									
Grouping and distinguish items #41 #19			x																									
Synchronism #38 #3											x																	
Visible transition #38 #3											x																	
Interaction state shared #38 #3											x																	
Predictability #38 #3											x																	
Adaptation at middle #38 #3 #9											x																	
Design Input Output #38 #3											x																	
Comments #39 #3											x																	
Prevention of errors #35 #3											x																	
Support for the user to recognize, diagnose and recover errors. #35 #3 #4 #7 #11 #14 #15 #19 #24 #27																	x											
Error Treatment #39 #3																	x											
Error prevention #4 #7 #14 #15 #19 #24 #27																	x											
Quality of Error Messages #41 #19																	x											
Number of Errors #6 #28 #30																	x											
Error protection #41 #19 #36 #4 #7																	x											
Correction of Errors #41 #19																	x											
Objective#8 #12																		x										
Structure #8 #12																		x										
Brevity #41 #19																										x		
Community #8 #12																											x	
Sophistication #8 #12																											x	
Symmetry #1				x																								
Balance #1				x																								
Contrast #1				x																								
Interface/Layout #23 #43				x																								
Functionality #23				x																								
Quality Content #8 #12 #22 #23				x																								
Coordination #23				x																								
Pattern #35 #3 #4 #7 #11 #14 #15 #19 #24 #27 #25 #36				x																								
Aesthetic design #35 #3 #36 #4 #7 #11 #14 #15 #19 #24				x																								
Textual information #25 #28 #43				x																								
Colors #28 #43				x																								
Scope #30				x																								
Importance code #41 #19				x																								
Emotion #8 #12				x																								
Promotion #8 #12				x																								
Compatibility #41 #19				x																								
Commands #43				x																								
Message #43				x																								
Windows #43				x																								
Font #43				x																								
Animations and Transitions #43				x																								
Graphic elements #43				x																								
Sound #43				x																								
Components #43				x																								
Icons #28 #43				x																								
Devices #43										x																		

Do I Really Have to Accept Smart Fridges?

An empirical study

Bushra Alolayan
Department of Computer Science
Aberystwyth University
Aberystwyth, Wales
bua3@aber.ac.uk

***Abstract*— Smart fridges have not taken off as rapidly as their creators would have wished. This raises questions about user attitudes towards the smart fridge. A small-scale empirical study, comprising 17 individual semi-structured interviews, results in identification key factors influencing user acceptance of the smart fridge. This leads to a new smart fridge acceptance model (SFAM). The SFAM modifies the well-known Technology Acceptance Model (TAM) by adding social influence, technology anxiety and cost to the factors included in the original Technology Acceptance Model. The findings indicate that participants are willing to use such an innovation if it already exists.**

Keywords-technology acceptance model; smart fridge; ubiqioutus computing; acceptance factors

I. Introduction

Ethnographic studies show that a huge part of people's social life occurs in the kitchen [19]. The kitchen is a place where you can combine vital communication and social purposes [30], a vast amount of time is spent there and many people regularly attach notes and other information to the fridge. Consequently, researchers from all over the world are working to enhance nutrition and dietary support [12], cooking, recipe planning [14] and communications [4].

This study is part of a three-year PhD project. The aim of this paper is to identify factors that will influence smart fridge acceptance based on the Technology Acceptance Model (TAM) [8] and provides useful services to users in the household. To do this, we performed the following:

- Reviewed the literature on technology acceptance.
- Identified the factors influencing the smart fridge based on reviewing the literature review and empirical investigation.
- Extended the Technology Acceptance Model based on the empirical findings and therefore proposed associated the factors with related variables.

New factors that influence the acceptance of the smart fridge have been identified empirically by following these three steps. This results in an improved acceptance model.

This paper is organised as follows. Section II presents an overview of the smart fridge and technology acceptance. In Section III, we empirically investigated set of factors influencing smart fridge acceptance. In Section IV and V, the results of interviews and discussion are presented, and finally, Section VI summarises the conclusions and future work.

II. Literature review

A. Smart Fridges

The smart Fridge is a smart home appliance that senses items placed inside the fridge and keeps track of stock through bar code or Radio-Frequency Identification (RFID) tags [26]; it also provides users with extra information about their products, their nutritional facts and consumption history.

The smart fridge has frequently been used as a prototypical instance of the Internet of Things. The smart fridge is a new concept; it consists of a large, flat surface that is fitted with a touch screen to allow interaction with users.

One function of the smart fridge is to offer a shopping assistant that helps the household to decide what food needs to be purchased. It, also, offers a meal planner that plans meals to be consumed and determines the necessary grocery items necessary for the preparation of those meals [13]. Lundberg [17] designed "the Snatcher Catcher", an interactive fridge that keeps track of the items in it using a camera. The results indicated that people showed a desire to own this fridge.

Today, the fridge also offers a facility to save energy; The Department of Energy and Climate Change in the UK published a report saying that using a smart fridge could save around 2 million tons of carbon dioxide every year and £222 million in energy savings [1]. Bigler [5] integrated a smart fridge into a Demand Site Management (DSM) network to reduce power peaks by more than 25% and to average out the overall energy consumption in order to improve grid utilization. Moreover, smart fridge specialists in "RLtec" have revealed that they are planning to install their demand management smart grid technology in smart fridges. This technology can be used to balance the power supply and demand [24].

According to Kuniavsky [16], "the fridge favoured technological unification over the user experience in three ways": lack of functional focus, value for money and ignorance of differences between the life cycles of consumer electronics and those of appliances. He argued that the reason behind lack of commercial success for the smart

 ISBN: 978-1-61208-325-4

fridge is that multiplying functions, such as support for general household management, distracts from the fridge's main purpose of keeping food fresh and offers no clear advantage over, for example, buying a tablet computer. In other words, the smart fridge should perform something that a regular fridge combined with a current computer cannot perform.

B. *Technology Acceptance*

User acceptance is crucial to the success of new technologies, but it is difficult to predict. Information system researchers have established numerous theoretical frameworks to understand why and how people accept a new technology. Some frameworks emphasize user acceptance of technology by using dependent variables such as intention or usage, while other frameworks emphasize implementation success at the organizational level. User acceptance refers to "the willingness of the user group to employ information technology for tasks the technology is designed to support" [9].

The Technology Acceptance Model (TAM) has gained a lot of attention from researchers and experts for many years. Hence, it is the most commonly adopted and effective model at determining the reasons future users accept or reject a specific information technology; it is also capable of adaptation to all attitudes in different contexts [11]. TAM assumes that primary determinants are perceived usefulness and perceived ease of use for attitudes towards using a particular technology [8]. Figure 1 shows the original TAM.

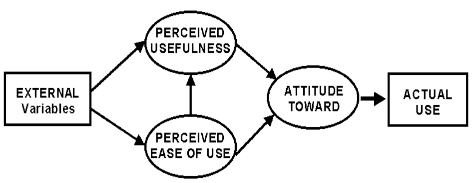


Figure 1. The original Technology Acceptance Model (TAM) [8].

Several empirical assessments of TAM recommend that perceived usefulness has constantly been the most important determinant of intention to use. Perceived usefulness is also affected by perceived ease of use.

The purpose of TAM is to clarify and investigate users' perception of a given system and to facilitate design changes before users have experienced a particular system. Moreover, the aim is to find out the effect of external variables on attitudes and intentions. TAM is an improvement on the Theory of Reasoned Action (TRA) model [2][10], which is influenced by beliefs, behavioral intentions and actual behavior.

The TAM proposed a set of factors that are important in determining user attitude towards accepting a technological innovation [8][18]. The Technology Acceptance Model factors are perceived usefulness, perceived ease of use, attitudes towards use are defined as "the user's desirability of his or her using the system" [18], actual use (AU) and external variables, such as demographic variables.

Venkatesh and Davis [32] proposed an extended TAM called TAM2, which involves social influence and cognitive instrumental variables into the original TAM. The social influence variable includes subjective norm and image, whereas cognitive includes job relevance, output quality and result demonstrability.

Venkatesh [33] extended TAM and proposed the Unified Theory of Acceptance and Use of Technology (UTAUT), which attempt to explain user intentions to use and usage behavior towards a new technology. UTAUT consists of seven constructs: performance expectancy, effort expectancy, social influence, facilitating conditions, moderating factors, behavioral intention and use behavior. The moderating factors are gender, age, experience, and voluntariness.

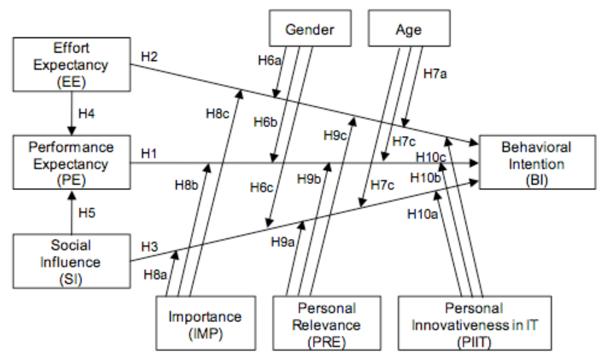


Figure 2. Smart kitchen acceptance model [21].

In a similar study, Rothensee [26] found that smart fridges are useful, easy to use and users will buy them if they are available on the market. He modified the TAM model with three main factors: perceived usefulness (PU), perceived ease of use (PEU) and affective attitude (A) towards behavioral intention. Also, moderators (gender, technological competence and sense of presence) have, according to the author, no significant impact on the model. Mayer [21] adopted the UTAUT model to analyze the user acceptance of smart products in smart home environment. Figure 2 shows the smart kitchen acceptance model.

TABLE I. FACTORS INFLUENCING SMART FRIDGE ACCEPTANCE BASED ON THE LITERATURE REVIEW.

Factors	TAM	UTAUT	Rotheness	Mayer et al.
SI	No	Yes	No	Yes
BI	Yes	Yes	Yes	Yes
PEU	Yes	Yes	Yes	No
PU	Yes	No	Yes	No
DF	Yes	No	No	No
A	No	No	Yes	No
E	No	Yes	No	No
PE	No	Yes	No	No
FC	No	Yes	No	No

User acceptance researchers think that more factors should be added to the TAM model or integrated with other acceptance models in order to improve its explanatory effect [20]. On the other hand, it has been found that TAM is the best model to explain users' attitudes toward new technology [3].

III. EMPIRICAL INVESTIGATION

The aim of this section is to present an empirical analysis that identifies the set of factors influencing the smart fridge.

A. Procedure

The initial literature review focused on acceptance models used in fields related to new technology, smart kitchens and smart homes. Based on those models, as well as semi-structured individual interviews, user acceptance factors in the smart fridge were identified using the Technology Acceptance Model because it is applicable in the context of innovative technology.

As the aim of the investigation was to gain a better understanding of the factors influencing acceptance of smart fridges rather than to test hypotheses, we adopted a qualitative approach.

The individual interviews were all semi-structured, consisting of mainly open-ended questions. According to Yin [35], “interviews are one of the most important sources of case study information.”

The interviews took about 30 minutes. Topics for discussion emphasized evaluation of the smart fridge system. In addition, quota-sampling technique used to identify willingness to participate.

The interviews conducted for this study consisted of questions about the interviewees’ technology background and demography, their attitudes towards the smart fridge as well as their attitudes towards smart kitchens in general. Also, the interviewees were asked about their lifestyle, eating habits, experiences and opinions regarding online shopping.

After obtaining ethical approval, email invitations were sent to the subjects to solicit voluntary participants. A link of a YouTube video about smart fridges and its features was sent with the invitation email. Participants watched the YouTube video before attending the interviews, which took place in a quiet environment to aid audio recording.

B. Participants

Subjects were students and staff members from different departments in Aberystwyth University. Altogether, 17 interviews were carried out from June 2013 until July 2013. Participants were restricted to over the age of 18 (females = 9) in order to cover and involve very diverse age groups in the study. Participants were recruited via courses and mailing lists at Aberystwyth University. Of our participants, 12 were students and 5 were self-employed. Table II provides a summary of the participants. In interview studies, the number of interviews included is normally recommended to be between 5 and 25 [31]. However, the number depends on the available time and resources for a particular study.

C. Data Analysis

All of the interviews were audio-recorded using a digital recorder and then transcribed for data analysis according to Grounded Theory [29]. These transcripts were imported into Dedoose web-based data analysis software for coding and analysis. We started with codes that had already been identified by previous literature and then added new codes from the interviews.

The code list included: attitude towards the smart fridge, risk, cost, performance risk, complex, enjoyment, privacy, social influence, memory loss, independence, usefulness and ease of use. This final code list covers two categories that are thought to impact users’ attitudes to smart fridge: Technical factors and social factors.

The second step of the interview analysis was identifying attitudes towards the smart fridge, which were either positive or negative. Out of the 17 volunteer interviewees, only 4 could be described as negative, 12 could be described as positive and the final interviewee could be described as moderate.

The final step of data analysis was to focus on and analyze participants with either positive or negative attitudes. The participant with moderate attitude was not selected for further analysis.

TABLE II. OVERVIEW OF THE INTERVIEWEES.

No.	Sex	Occupation	Age	Social Status
1	F	Postgraduate Student	25	Single
2	F	Lecturer	46	Married
3	M	Undergraduate student	20	Single
4	M	Postgraduate Student	26	Single
5	M	Postgraduate Student	27	Single
6	F	Secretary	47	Married
7	F	Postgraduate Student	35	Married
8	F	Postgraduate Student	27	Single
9	F	Postgraduate Student	26	Single
10	M	Lecturer	49	Married
11	F	Postgraduate Student	25	Single
12	M	Postgraduate Student	25	Single
13	F	Postgraduate Student	26	Single
14	M	Postgraduate Student	28	Single
15	F	Postgraduate Student	25	Single
16	M	Lecturer	45	Married
17	M	Lecturer	47	Married

IV. RESULTS

The qualitative analysis suggested that users’ intention to accept the smart fridge is affected by many factors. Based on the data analyses, we grouped these factors into two categories: technological and social factors.

Results were analyzed with selected quotations and not all of the participants are quoted for each part. The findings are based on our analysis of all of the interviews according to common practice in qualitative studies.

None of the participants had any experience with either smart fridges or smart kitchens. However, it appears that participants accepted the idea of smart fridges and showed a high willingness to use them. On the other hand, there were numerous aspects identified that make the implementation and acceptance of smart fridge measures more difficult.

The factors mentioned by the participants are described below. All names of participants have been changed.

A. Social Factors

1) Cost

Participants agreed that the cost of a smart fridge would prevent them from using it; the current cost of a smart fridge is around £3000 [25]. Kuniavsky [16] argued that the market price for each of these smart fridges was more expensive than the combination of its technologies. Prices this high belong to luxury goods, where the price of the goods is not based on the functionality. Thus, the price of the smart fridge should be based on its functionality and its ability to deliver a better user experience.

From individual interviews, there were mixed reactions towards the smart fridge price, 83% of the participants were willing to spend up to £600 on the smart fridge. Some participants agreed that the price would reduce in the future, just like other smart gadgets have, and that every household would own a smart fridge in the future because *"you have got to have it"*. (Mark)

John compared smart fridges with LED TVs: *"At the moment I think it is going to be a significant issue. However, years ago an LED TV cost around £3000-4000; now it is for £300. So, what we have to assume at some point is that the price will reduce. Because the benefits are less for me than for the next 20 years I won't spend so much money."*

Moreover, participants argued about the risk if the smart fridge system goes wrong; they will end up with a basic fridge costing much less than the smart one. Thus, it is relative to the perceived usefulness of the system. Kate explained: *"Price is a serious factor. It is a big investment, you know, if part of it goes wrong, if the computer part goes wrong, you basically just have a standard fridge again which you can get for a couple of quid."*

2) Technology Anxiety

Technology anxiety is defined as "the fear and apprehension people feel when considering use of or actually using technology-related tools" [7][22][28].

Most university students and staff members are comfortable with using computers and technology. The results show that all of the participants have a good knowledge of using computers. However, even when participants can see the benefits of using the smart fridge, they may avoid it if they are not comfortable with using such a new technology.

Other participants argued that they feel anxious using new technology, especially when initially interacting with it, but the anxiety disappeared as they familiarized themselves with the functionality, for example, as David explains: *"I don't have any problems using new technology but I feel anxious maybe for the first 10 minutes; after that I think I get used to it. I don't give up that easily; I just try to work it out."*

3) Social Influence

Moore and Benbasat [23] defined it as "the extent to which use of an innovation is perceived as enhancement of one's status in a social system". Davis [8] believed that, in some cases, people might use a system to comply with others' mandates rather than their own feeling and beliefs. Mark explains: *"I think at some point in the future everybody will buy it because everybody has to own it. Like the smart phone - nowadays everyone owns one because everybody does!"*

In addition, some participants indicate that that behavioral intention to use a smart fridge is influenced by their friends, colleagues and the community. For example, as Susanne explains: *"If I heard any good reviews about the smart fridge from friends I will definitely buy it."*

B. Technical Factors

1) Perceived Usefulness

Davis [8] in his original TAM defined it as "the degree to which an individual believes that using a particular system would enhance his/her job performance."

The results of the interviews indicate that participants all acknowledged the system's usefulness in terms of its potential benefits to save food and make their life much easier. For example, most participants said that they think the smart fridge is *"very useful."*

The smart fridge is useful for elderly people as they experience a decrease in memory capacity. Memory loss for older adults plays an important role in their preference for the smart fridge. Smart fridges will help them to generate shopping lists automatically, as well as know what is cooling inside the fridge without depending on their memory. Steve explains: *"I think that very often we forget what we have in the fridge we go shopping and then you buy the same thing again."*

Also, Steve added: *"I have aging parents so we recognize that at some point it is going to be difficult to remember everything in the fridge, so I see that the smart fridge is quite a useful facility, as we get older."*

The smart fridge may increase the effectiveness and efficiency of the participant's life by offering many features such as grocery shopping and recipe suggestions. Rose explains: "*It will help me when I am shopping for food. The fridge will help me to not buy unnecessary food.*"

2) Perceived Ease of Use

This is defined as "the degree to which an individual believes that using a particular system would be free of physical and mental efforts." [8]

The majority of participants spoke about the level of user-friendliness of the smart fridge. Since most of the participants have good experience with computers, it was expected that they would be comfortable with using a smart fridge in the future to track items and check consumption history; however, they think that interacting with this technology will be easy. If things are not perceived as easy, participants will simply not use them. Hannah explains: *"If it's complicated I don't think I'm going to enjoy using it."*

Good interface design will also increase perceived ease of use, thereby attracting more users to the smart fridge *"I think the GUI itself will be important. Also, it has to be easy to control"* (David). Moreover, the convenience of the smart fridge and its ability to make grocery orders and deliver them to the household reduces travelling, thus increasing the well-being and independence as well as increasing the ease of everyday activities. For example, Lauren explains: *"I needed it the most when my children were young and it was not easy to take the baby to the store."*

It is, therefore, assumed that when participants perceive the smart fridge as easy to use, they will be more likely to accept it.

V. DISCUSSION

The findings suggest that motivation for accepting the smart fridge is affected by numerous factors that have been categorized as follows: social and technological factors. Five main factors have been identified as important for smart fridge technology acceptance: perceived usefulness, perceived ease of use, social influence, technology anxiety and cost.

The interviews confirmed the major elements such as perceived usefulness and perceived ease of use in affecting users' motivation to accept the smart fridge. Combing the findings from the literature review and the results of the interviews, an enriched research model, the Smart Fridge Acceptance Model (SFAM), is proposed with three additional constructs added: social influence, technology anxiety and cost (Figure 3).

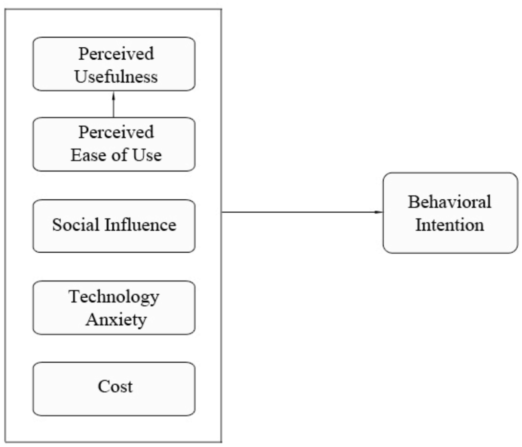


Figure 3. Smart Fridge Acceptance Model (SFAM).

Perceived usefulness and ease of use are considered the most important factors in smart fridge acceptance. The easier the system is to use, the more users will find it helpful, which leads to an increased intention to use a smart fridge in the future.

In terms of perceived usefulness, participants consider particularly useful tracking items and expiry dates. This can also be very useful for elderly people as they experience decrease in memory capacity.

In the area of perceived ease of use, the results reveal that participants think that the smart fridge will be easy to use, which confirms findings from existing technology acceptance literature. Each participant indicated familiarity with computers therefore makes smart fridge easy to use.

Participants with satisfying computer experience and low technology anxiety were more likely to use the smart fridge in the future and more likely to spread positive word-of-mouth. Therefore, increasing computer experience will reduce technology anxiety.

Participants are generally uncomfortable with uncertainty and therefore consult friends, family members and the social network on their adoption decisions [6]. Moreover, participants regard this kind of new technology as symbolic of wealth to enhance their sense of self-importance [27].

The only limitation observed from the interviews was that participants are not willing to own a smart fridge because of its high price. That is in line with the existing research of Kim [15] and Wu [34], who found a negative impact of the cost on technology acceptance. Participants also expressed willingness to try out the smart fridge if it was readily available.

Overall, the majority of the findings support the findings from existing literature and related technology acceptance research.

VI. CONCLUSION AND FUTURE WORK

This paper investigated the factors influencing the acceptance of the smart fridge. We studied existing Technology Acceptance Models and extended a set of factors that could influence smart fridge acceptance. These factors were identified by studying the literature reviews and then validated the qualitative interviews. The original TAM was the best user acceptance model to analyze smart fridge acceptance; however, the TAM needs to be extended to identify smart fridge factors. The results show that perceived usefulness, perceived ease of use, social influence, technology anxiety and cost are the most important factors for smart fridge acceptance.

The advantage of this research is that its focus is on the user, who is the key to acceptance of a new technology; it was also based on the original Technology Acceptance Model, which has been validated in several studies. The main contribution of this study is the domestication theory that was used for proposing the SFAM.

Future research will involve measuring each factor using multi-item scales, as well as further empirical investigation of the model using a smart fridge prototype.

ACKNOWLEDGMENT

The author would like to thank participants, the reviewers of this paper and Edel Sherratt for their cooperation in this study.

REFERENCES

[1] J. Adetunji, "UK set to trial 'smart fridges'", The Guardian, Available online at www.theguardian.com, 2008.

[2] I. Ajzen, and M. Fishbein, Understanding attitudes and predicting social behaviour, Englewood Cliffs, NJ: Prentice-Hall, 1989.

[3] K. Alshare, E. Grandon, and D. Miller, Antecedents of computer technology usage: considerations of the technology acceptance model in the academic environment, Journal of Computing Sciences in Colleges, vol. 19, April 2004, pp.164-180.

[4] J. Bauer, K. Streefkerk, and R. R. Varick, Fridgets: digital refrigerator magnets, In: CHI 2005 Student Design Competition, Portland, 2005, pp. 2060–2064.

[5] T. Bigler, G. Gaderer, P. Loschmidt, and T. Sauter, SmartFridge: Demand Side Management for the device level, Emerging Technologies & Factory Automation (ETFA), 2011 IEEE 16th Conference on IEEE. Toulouse, pp. 1-8, 2011.

[6] M. E. Burkhardt, and D. J. Brass, Changing patterns or patterns of change: The effects of a change in technology on

social network structure and power, Administrative science quarterly, vol. 35, no. 1, March 1990, pp. 104-127.

[7] M. A. Cambre, and D. L. Cook, Computer anxiety: Definition, measurement, and correlates, Journal of Educational Computing Research, vol. 1, 1985, pp. 37-54.

[8] F. D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS quarterly, 1989, pp. 319-340.

[9] A. Dillon, and M. G. Morris, User acceptance of new information technology: theories and models, Annual review of information science and technology, vol. 31, 1996, pp. 3-32.

[10] M. Fishbein, and I. Ajzen, Belief, attitude, intention and behavior: An introduction to theory and research. Reading, MA: Addison-Wesley, 1975.

[11] L. Gentry, and R. Calantone, A comparison of three models to explain shopbot use on the web, Psychology & Marketing, vol. 19, pp. 945-956, November 2002.

[12] V. Hanson-Smith, D. Wimalasuriya, and A. Fortier, NutriStat: tracking young child nutrition. CHI'06 extended abstracts on Human factors in computing systems. ACM, 2006, pp. 1831-1836.

[13] M. Janzen, and Y. Xiang, Probabilistic reasoning for meal planning in intelligent fridges, Advances in Artificial Intelligence. Springer, 2003.

[14] W. Ju, R. Hurwitz, T. Judd, and B. Lee, CounterActive: an interactive cookbook for the kitchen counter, CHI'01 extended abstracts on Human factors in computing systems, ACM, 2001, pp. 269-270.

[15] H. W. Kim, H. C. Chan, and S. Gupta, Value-based adoption of mobile internet: an empirical investigation, Decision Support Systems, vol. 43, February 2007, pp. 111-126.

[16] M. Kuniavsky, Smart things: ubiquitous computing user experience design, Morgan Kaufmann, 2010.

[17] J. Lundberg, A. Ibrahim, D. Jönsson, S. Lindquist, and P. Qvarfordt, The snatcher catcher: an interactive refrigerator, Proceedings of the second Nordic conference on Human-computer interaction. ACM, 2002, pp. 209-212.

[18] Y. Malhotra, D. F. Galletta, Extending the technology acceptance model to account for social influence: theoretical bases and empirical validation. System Sciences, 1999. HICSS-32, Proceedings of the 32nd Annual Hawaii International Conference on IEEE, 1999, pp. 14.

[19] M. Mateas, T. Salvador, J. Scholtz, and D. Sorensen, Engineering ethnography in the home, Conference companion on Human factors in computing systems, ACM, 1996, pp. 283-284.

[20] K. Mathieson, Predicting user intentions: comparing the technology acceptance model with the theory of planned behavior, Information systems research, vol. 2, September 1991, pp. 173-191.

[21] P. Mayer, D. Volland, F. Thiesse, and E. Fleisch, User Acceptance of 'Smart Products': An Empirical Investigation, In WI 2011 Proceedings, vol. 2, 2011, pp. 1063 – 1072.

[22] M. L. Meuter, A. L. Ostrom, M. J. Bitner, and R. Roundtree, The influence of technology anxiety on consumer use and experiences with self-service technologies, Journal of Business Research, vol. 56, 2003, pp. 899-906.

[23] G. C. Moore, and I. Benbasat, Development of an instrument to measure the perceptions of adopting an information technology innovation. Information systems research, vol. 2, 1991, pp. 192-222.

[24] J. Murray, Smart fridge specialist sets sights on supermarket chillers. Business Green, April 2010.

[25] Reuters. Samsung's T9000 smart refrigerator runs on Android, includes apps like Evernote and Epicurious. NDTV Gadgets, January 2013.

[26] M. Rothensee, User acceptance of the intelligent fridge: empirical results from a simulation, The Internet of Things, 2008, pp. 123-139.

[27] S. Sarker, and J. D. Wells, Understanding mobile handheld device use and adoption, Communications of the ACM, vol. 46, 2003, pp. 35-40.

[28] C. R. Scott, and S. C. Rockwell, The effect of communication, writing, and technology apprehension on likelihood to use new communication technologies, Communication education, vol. 46, 1997, pp. 44-62.

[29] A.L. Strauss, J. Corbin, and S. Niewiarra, Grounded theory: Grundlagen qualitativer sozialforschung, Beltz, PsychologieVerlagsUnion Weinheim, 1996.

[30] L. Swan, and A. S. Taylor, Notes on fridge surfaces, ACM, 2005, pp. 1813-1816.

[31] J. Vaivio, Interviews–Learning the Craft of Qualitative Research Interviewing. European Accounting Review, vol. 21, 2012, pp. 186-189.

[32] V. Venkatesh, and F. D. Davis, A theoretical extension of the technology acceptance model: four longitudinal field studies, Management science, vol. 46, 2000, pp. 186-204.

[33] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, User acceptance of information technology: Toward a unified view, MIS quarterly, 2003, pp. 425-478.

[34] C. S. Wu, F. F. Cheng, D. C. Yen, and Y. W. Huang, User acceptance of wireless technology in organizations: a comparison of alternative models, Computer Standards & Interfaces, vol. 33, 2011, pp. 50-58.

[35] R. K. Yin, Applications of case study research (applied social research methods series volume 34). London: Sage, 1993.

Nonintrusive Multimodal Attention Detection

Hugo Jiawei Sun, Michael Xuelin Huang, Grace Ngai and Stephen Chi Fai Chan
Department of Computing
The Hong Kong Polytechnic University
Hong Kong S.A.R.
{hugo.sun, michael.xl.huang}@connect.polyu.hk, {grace.ngai, stephen.c.chan}@comp.polyu.edu.hk

***Abstract*— With the increasing deployment of computers in a wide variety of applications, the ability to detect the user's attention, or engagement, is becoming more important as a key piece of contextual information in building effective interactive systems. For instance, one can imagine that a system that is aware of whether the user is attending to it would be able to adapt itself better to the user activities to enhance productivity. The ability to detect attention would also be useful for system analysis in designing and building better systems. However, much previous work in attention detection is either obtrusive or imposes demanding constraints on the context and the participants. In addition, most approaches rely on uni-modal signals, which are often limited in availability and stability. This paper attempts to address these two major limitations through a noninvasive multimodal solution, which allows participants to work naturally without interference. The solution makes use of common off-the-shelf items that could reasonably be expected of any computing environment and does not rely on expensive and tailor-made equipment. Using a three-class attention state setting, it achieves average accuracy rates of 59.63% to 77.81%; the best result being 77.81% for a general searching task, which shows 11.9% improvement over the baseline. We also analyze and discuss the contribution by individual features to different models.**

Keywords-Affective computing; keystroke dynamics; facial expression; multimodal recognition; attention detection

I. INTRODUCTION

Human attention is strongly correlated with a person's efficiency in both working and studying. Considering the pervasiveness of computers today, it makes sense to provide better interaction between users and computers by taking into account contextual information, such as location, expertise, and preference, in the development of applications [1]. However, one of the most important factors in Human Computer Interaction (HCI), the mental state of a user, has been largely ignored in real world applications.

Affective computing [2] and psychophysiology represent the most important areas in HCI towards the understanding and utilization of human mental states, and there has been some work on the automatic detection and recognition of human mental states by the computer [1, 3]. Among the various mental states, it is known that attention is one of the most significant and indicative affects that affects a user's efficiency, productivity and even creativity [4].

Recent advancements in technology have enabled more precise estimation of a human's psychological states in the form of physiological signals, which formerly could only be acquired through the use of intrusive and sometimes even invasive medical sensors. These products, such as the NeuroSky Mindset [5] and the Emotiv EPOC [6], use a headset with dry sensors to detect electroencephalogram (EEG) signals, and report the alpha and beta waves. This is a large step forward in term of convenience compared to the traditional clinical use of EEG, but these devices are still intrusive as they require a user to buy and wear them. Therefore, we believe that a truly nonintrusive approach in obtaining the attention, concentration, and engagement levels of a user is necessary and would be a contribution to the field. Previous work has explored vision-based approaches that estimate the concentration level by analyzing the facial expression of a user [7], and keystroke dynamics and mouse activities have been used to classify a user's mental state of engagement and boredom [8]. These research works focus on attention detection through analyzing signals from a single modality.

We believe that for real world applications, uni-modal attention detection approaches may be easily affected by the contextual factors and individual differences. To address this problem, we propose a multi-modal approach to detect attention level in a nonintrusive manner. Our challenge lies in the feature representation and selection of modalities. Most of the features we adopted have been proven effective in state-of-the-art research, while the rest are introduced specifically for this work. Our contributions include the introduction of useful features in attention recognition, the analysis of contributing factors of different features from distinct modalities, the evaluation of our proposed solution along the feature and temporal dimensions, and the identification of important features that are closely related to human attention states.

The rest of the paper is organized as follows. Section II first presents an overview of research work in related fields, including facial expression and keystroke application. Section III describes the multimodal signal features in detail and illustrates the collection of ground truth after a short analysis on problems of existing solutions. In addition, it also illustrates the experimental setting and explains the choice of experiment tasks. Section IV follows with a performance analysis on proposed models. Results are presented and interpreted from both specific and general aspects. Based on the results, Section V discusses the findings and suggests some potential problems of the work.

Finally, the paper ends with an outline of future research direction and a brief summary in Section VI.

II. RELATED WORK

State-of-the-art affective computing research usually uses features from a wide range of areas: facial expression, vocal intonation, hand gesture, body posture, language, and physiology signals [1, 3]. Of the above, physiological signals are the most able to provide a precise measurement, but they are often intrusive, in the sense that experimental subjects are required to have different sensors, like Electrocardiogram (ECG), Electromyogram (EMG), and EEG, attached onto their bodies. On the contrary, nonintrusive approaches rely on observations of expression, gesture, posture, vocal intonation and language, which are captured by camera and microphone. Of these, relatively less attention has been paid to body posture and hand gesture, partially because it is more difficult to accurately infer affective states from these modalities, and also because there is some evidence that posture can only reflect the quantity (intensity) of the emotion, instead of its quality (category) [9]. More recently, keystroke dynamics have been investigated as a potential nonintrusive input for emotion detection [8], [10].

In this paper, we focus on detecting a user's attention level for various tasks that commonly occur in a workplace setting. Therefore, our approach requires us to use information that can be collected in a nonintrusive approach from equipment that is available readily in a standard desktop computer. Since similar approaches have been used to detect different dimensions and categories of emotions and cognitive mental states, we believe that it is also possible to detect a user's attention level based on these modalities in a nonintrusive manner.

A. Facial expression

Facial affect detection is one of the most popular approaches in affective computing. Since facial expressions are strongly correlated with emotions, especially basic emotions [11], most vision-based affect recognition research focuses on facial expression analysis [3], mostly to recognize the basic emotions [1, 3], or higher level affects (e.g., interest) and cognitive states (e.g., thinking). El Kaliouby et al. [7] use dynamic Bayesian networks to model and recognize six complex mental states, namely, agreeing, concentrating, disagreeing, interested, thinking and unsure. Lan et al. compare static and dynamic approaches to model and infer fatigue [12], taking into account environment factors (e.g., noise, temperature, and humidity), and personal physical states (e.g., sleep quality) and visual hints (e.g., head and eyelid movements). Ashraf et al. [13] and Littlewort et al. [14] attempt to infer pain from facial expressions. Subjects are required to act and feel pain by putting their hands into the icy water in the control experiments. In addition, differences in both appearance and timing between posed (simulated) and natural facial expressions have been observed.

B. Keystroke

Emotion detection based on keystroke dynamics has a relatively short history, despite its remarkable success in authentication [15], [16]. Vizer et al. [17] detect cognitive and physical stress based on keystroke dynamic. Besides traditional keyboard event features, their method also includes linguistic features like emotive word rate, verb rate, conjunction rate, etc. Epp et al. [10] create 2-level classifiers using keystrokes to recognize 15 emotional states, among which confidence, hesitance, nervousness, relaxation, sadness and tiredness accord at least 27% accuracy better than chance. The essential keystroke features used are timing patterns of single keystrokes, digraphs and trigraphs. Bixler et al. [8] investigated 14 emotion states, based on pausing behaviors, keystroke verbosity and timing, and found that only 6 of them occur with some regularity, among which engagement and boredom rank as the most frequent states.

Our proposed method distinguishes itself from the previous works in several essential aspects. First, for attention level detection, we apply multimodal features, including vision, keyboard, mouse, etc. This allows our method to take into account a wide context of the user's environment. Second, we distill the essential features into different modalities that contribute significantly in the attention level classification. Third, we attempt to classify attention levels according to different time intervals and compare their performances. Fourth, we generalize our approach and evaluate it across dissimilar working tasks performed by different subjects. Finally, we also propose a novel feature to quantify mouse movements.

III. METHODOLOGY

Our method approaches the problem from a nonintrusive and multimodal perspective. Since we are interested in real-use scenarios, our approach assumes only equipment and peripherals that would commonly be available on computers in an office setting, and signals that can feasibly be collected via such devices.

For ease of explanation, we classify our incoming signals into machine- and human-specific features. Machine-specific features include keystrokes, mouse activity, window layout, and so on, whereas human-specific features include facial expressions and head movement. As the various input devices collect data at different frequencies, the raw incoming data is aligned at the desired frequency via re-sampling to create a feature vector for every second. A nonparametric random forest machine learning algorithm then analyzes the features to identify indicative features.

A. Keyboard features

As one of the most often-used input devices in a work setting, the keyboard is one of the most indicative recorders of a user's behavior and activity.

Table I shows the different keyboard-related features used in our approach. These keyboard features make use of both statistical and grammatical information to characterize a user's activity. They include 1) counting of key presses and 2) time intervals between presses. For example, Feature KB,

TABLE I. KEYBOARD FEATURES

Code	Description
KB	The number of key presses per second
BS	The number of *Backspace* and *Delete* key presses per second
NWIS	The number of words since the most recently-finished sentence
NLIW	The number of letters since the most recently-finished word
VWIS	Word input rate of the last sentence
VLIW	Letter input rate of the last word
STD	Time interval since last *Space* key press
DTD	Time interval since the last ".", "?", and "!" press
KIDT	Keyboard idle time

as a general parameter reflecting overall keyboard activities, counts the total number of keyboard press each second, including letter, number, punctuation, and control keys. Feature BS measures the frequency that a user corrects a typo, which implies the user's level of awareness, and, indirectly, the level of attention. Features NWIS and NLIW are delimited by "space" and "punctuation", where "space" signifies the end of a word, while punctuation marks like period ("."), question mark ("?"), and exclamation mark ("!") indicate the end of a sentence in modern English. The assumption behind these two features is that generally, long and complex language processing requires more thinking, and these features imply language information that in turn would suggest human attention. Features STD and DTD record the time intervals between two spaces and punctuations, which are used for the calculation of word and sentence input rate together with features NWIS and NLIW. These keyboard features are highly useful especially for typing tasks.

In addition, feature KIDT is introduced to characterize "idle time" on the keyboard. This feature is activated when the machine has received no key press for 1 minute. There are no conventions that clearly define idle state (i.e., how many seconds of inactivity counts as an idle state?), so this 1-minute threshold is used as a starting point and subject to further investigation. When the keyboard enters the "idle" state, it resets all other keyboard statistical features, under the assumption that they are now outdated.

The data are further extracted into time frames of 5, 10 and 15 seconds as in previous approaches [8], [10], under the assumption that attention is a temporally changing state. Taking this information into consideration helps us to understand attention from a temporal aspect. In the experiments, these features are denoted with the extensions: -5, -10, and -15 (e.g., KB-5, KB-10, and KB-15).

B. *Mouse features*

Other than the keyboard, much of the rest of the user input takes place through the mouse. Mouse activities include click, move and wheel rotation. Table II gives a description for each of them.

TABLE II. MOUSE FEATURES

Code	Description
MC	The number of mouse click per second
MM	Mouse move by pixel per second
MR	Mouse wheel rotation angle per second
UM	Unnecessary mouse movement
MIDT	Mouse idle time

Intuitively, characterizing mouse activity might start with seeing how far the mouse has moved and how often it triggers a click. Feature MC counts the mouse click per second, while Feature MM captures the mouse move as a number of pixels. In addition to move and click, Feature MR, reflecting the rotation direction of mouse wheel, is in use when the user is browsing and reading. Since the program is able to detect direction of rotation (e.g., toward the monitor or backward to the user), this feature can reveal some interesting user behavior (e.g., disordered reading pattern).

Besides the aforementioned features, we also introduce the new Unnecessary Mouse Movement (UM) feature. The rationale behind this feature is based on the assumption that users moving the mouse with a clear target in mind will normally follow the shortest path from their current location to the target point, and an observation that, often, perhaps as a nervous habit, distracted users often move the mouse around somewhat randomly, without clicking (as this would not trigger an operating system event). Unnecessary Mouse Movement is defined by the difference between 1) the actual mouse movement path (the dotted line) and 2) the shortest distance (the solid line) between the current cursor position (B) and the position of the last mouse click (A) (**Figure 1**). This allows us to quantify random and aimless mouse movements, which may be potential indicators of a lack of user attention.

Similar to keyboard features, we extract the data by time frames of 5, 10 and 15 seconds to obtain a temporal view.

C. *Facial features*

From real-life experience, facial expressions, including head gestures, are good indicators of a person's attention. Among the various representations of face and head features, facial action units [11] have been proven to be valuable interpretations of facial expression in both psychology and computer vision [7], [14], [18]-[20]. They are the essential representation of a facial expression – it is possible to describe all facial expressions by combining different facial action units. Furthermore, facial action unit based expression analysis is relatively robust. Once the facial landmarks are accurately identified, the values of action units will not be influenced by the variation of environmental factors, such as

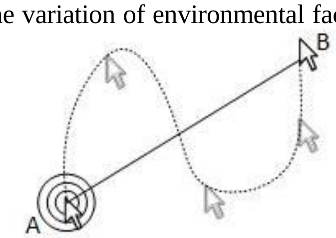


Figure 1: Unnecessary Mouse Movement (UM)

TABLE III. FACIAL FEATURES

Code	Description
FF1-FF4	Inner and outer eyebrow movement
FF5-6	Eyebrow movement
FF7-8	Eyelid movement
FF9	Upper eyelid movement
FF10-11	Lip corner puller or depressor
FF12	Lower lip depressor
FF13	Lip pucker
FF14	Lip stretcher
FF15	Lip funneler
FF16	Lip tightener
FF17	Lip separated
FF18-FF22	Head movement
FF24	Eye gaze

light intensity change and shadow movement.

In this paper, we apply Constrained Local Models (CLM) [21] to track 66 facial landmarks. The model is trained on the CMU Multi-PIE Face database [22], which contains over 750 thousand images from 337 individuals. The CLM optimization procedure iteratively adjusts the 2D and 3D landmarks and other global and local parameters.

Figure 2 shows the facial landmarks tracked by CLM and Table III presents the 24 facial features adopted in our work. The wired models indicate the 2D and 3D tracked facial landmarks, and the lengths of the green bars denote the intensities of the facial features. The facial features we used are similar to motion units [23], which describe facial movement; but are numeric in nature and represent both direction and intensity of the facial movements.

Similar to previous work [24], we represent the degree of a motion unit by the normalized distance between the corresponding feature points. To reduce the influence caused by head orientation, we use the 3D landmarks to calculate the features in the 3D space, which is more stable than calculating the distance features from the aligned face through the affine transformation. The five head movement features, "FF18-FF22", capture the head scaling, orientation (pitch, yaw and roll) and position (x-axis and y-axis translation). Since the CLM model we used does not track iris movement, we extract the iris based on the integral image [24] of the region covered by the eye landmarks. The darkest sub-region in the integral image is considered to be the iris. Feature "FF24" represents the distance proportion from the pupil center to eye corners.

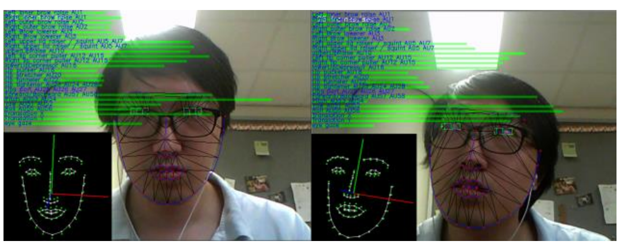


Figure 2 Illustration of CLM face tracking results of two facial expressions (left: attentive; right: non-attentive).

TABLE IV. MACHINE-SPECIFIC FEATURE

Code	Description
WS	Proportion of current active windows to desk top

D. Other features

In addition to the keyboard and mouse features, we introduced another machine-specific feature: active window size (Table IV). The active window size is the proportion of size of the window relative to the whole screen. Our hypothesis is that the bigger the window size, the less likely the user would be distracted by irrelevant information on the screen.

Our input features cover a diverse spectrum of the user's activity on the computer; however, there is one obvious feature that is omitted – that being the focus of the user's eye gaze on the screen. We had indeed considered inclusion of that feature, but it is very difficult to obtain that information accurately without the use of specialized eye-tracking equipment and/or requiring the use of wearable devices. Since this is not compatible with the assumptions and objective of our project, we made the decision not to include this as a feature.

E. Ground truth collection

In order to classify the incoming signals, the individual vectors need to be labeled with the "ground truth" for training and testing.

Previous work has pioneered the use of the Neurosky Mindset device [27] as an off-the-shelf EEG device to provide the ground truth in an objective manner for attention detection [26]. The Mindset collects brainwave data (i.e., alpha, beta, and theta waves) [27] and also processes the raw brainwave data into integer values ranging from 0 to 100, which indicate the level of "concentration" of the user. The accuracy of the Mindset's attention detection algorithm has been verified in previous work [27], which measured a positive correlation between the Mindset-measured and user-self-reported results. Given these results and the fact that we wish to detect attention as it relates to brain activity level, e.g., the intensity of mental "focus" that occurs during intense concentration [28], we believe that the Mindset is reliable as a measure of ground truth for our work.

The MindSet's attention detection algorithm returns an integer that indicates a user's attention level. The MindSet Instruction Manual [28] provides an interpretation that classifies the attention level into five equally divided classes of incremental attention level. We further simplify this into a three-class model dividing the range uniformly across the three classes to produce the three levels: *non-attentive*, *neutral*, and *attentive*.

F. Machine Learning Classification

The data collected for this project includes multiple signals from various channels. These features may contribute differently, depending upon the user's context or his/her task. For example, we would hypothesize that the keyboard-based activities would be more indicative during typing tasks, but

 ISBN: 978-1-61208-325-4

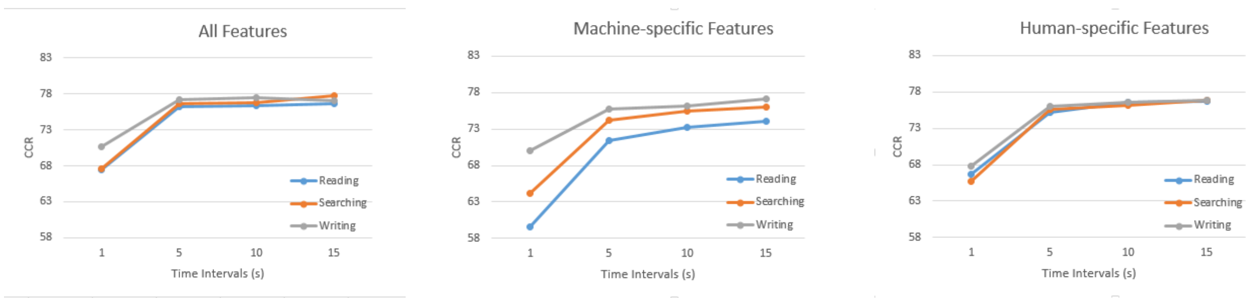


Figure 3: Performance Changes as a Function of Time Interval

would contribute far less when the user is web browsing. Therefore, our machine learning algorithm has to be able to identify useful features among large numbers of signals and to avoid overfitting at the same time.

We use the Random Forests [29] machine learning algorithm for this work. The random forest algorithm constructs multiple decision tree classifiers and aggregates the result using the mode of the output results of the individual classifiers. It is known to have several advantages: 1) it yields better accuracy than many current algorithms; 2) it can handle large number of variables; and 3) it weights importance of different features; 4) it is known to be fairly robust to overfitting; and finally, 5) it does not require unified data types in the same feature vector.

G. *Experimental Tasks and Setup*

A total of 10 participants were recruited for our experiments (5 male and 5 female, aged 22-29, 3 undergraduates and 7 graduates). All participants were proficient in using a computer running Windows 7 and read and write English fluently. They were first asked to search for academic papers relevant to their research or coursework for 30 minutes. They then spent another 30 minutes reading papers. Finally, in the last 30 minutes, they did some academic writing. The materials involved in the three tasks may or may not be relevant to each other.

The three tasks were chosen because they are simplified versions of commonly-performed tasks in the office or lab space, and thus provide a realistic picture of user interaction. The tasks also represent three common modes of interactions with the computer: input-oriented, output-oriented, and hybrid, which triggers or relies on different interaction patterns: in the reading task, neither keyboard nor mouse is active, while in writing, the keyboard produces the main source of signals. Finally, the searching task expects both keyboard and mouse activities.

The experimental setup was kept as close to that of a normal workplace as possible. With the exception of the Mindset, which was connected to the computer via a Bluetooth link, the rest of the experimental setup consists of equipment that would commonly be found in an office or lab setting: a desktop computer running Microsoft Windows 7 with keyboard, mouse, and webcam.

H. *Data logging and preprocessing*

The preprocessing procedure transforms the computer interaction log into the set of previously-mentioned features. The programs use windows hook to collect data and produce event-based logs. A new entry is appended to the log whenever there is a key press or mouse activity on the computer. At the same time, the Mindset device returns a temporal log. The raw data is then preprocessed into features suitable for classification. Human-specific features from the video stream and machine-specific features (keyboard, mouse, etc.) from the computer log were extracted according to the re-sampling frequency, and then were aligned with the temporal readings from the Mindset.

IV. RESULTS

A total amount of 54000 seconds of responses were gathered, each participant contributing about 5400 seconds. The data from each task was preprocessed into four sets with different lengths of time frames: 5, 10 and 15 seconds, as previously mentioned. Following previous work [17], we evaluated the results using 10-fold cross-validation. The results were also compared across different feature combinations, i.e., machine-specific features alone, human-specific features alone, and fusion of the two. We use accuracy, i.e., Correctly Classified Rate (CCR) and Kappa statistics [30] to study the classification performance.

One observation that was made early on in our experiments was that the raw data collected follows a normal distribution. Given this, our three-class model for the ground truth places more than half of the data into the *neutral* class. This uneven distribution obviously will affect the evaluation benchmark, so we also provide a comparison between our model and the “most-frequent class” baseline to provide a fairer picture.

A. *Performance of specific models*

Table V shows our evaluation results. Our method constructs individual models for each user in each task. The mean and standard deviation of the performance by CCR are presented, and the best value along each row is bolded. The reported performances here represent the average performances over all subjects.

The results, achieved from three-class models (*attentive*, *neutral*, and *non-attentive*), are generally appreciable. The best result reported is 77.81%, when the searching task is performed, with a moving window of 15 seconds. It is 11.90% better than the benchmark 65.91%.

An inspection of our results shows that fusion of both human-specific and machine-specific features achieves better results. Almost all (11 out of 12) of the fused, multimodal models achieve the best performance for each sample set.

This bears out our hypothesis that a multimodal approach that integrates both machine-specific and human-specific features performs better than using either of the two alone.

Besides feature combination, we can also analyze the performance along the dimension of window size. As the time interval increases, the classification accuracy increases as well. For example, on the reading task, using only machine-specific features, changing the window size from 1 to 15 seconds gives us an accuracy gain of 14.46%, which outranks the improvement achieved using feature combination. Generally speaking, as the window size increases from 1 to 5 seconds, the performance increases significantly, and flattens out after that (Figure 3). In our research, a moving window with time interval of 15 second yields the best result. This is consistent with previous work [8], [10], [27], which also use a 15-second time interval in data preprocessing, albeit for uni-modal models.

B. *Performance of general models*

Our results so far produce a user-specific, task-specific model. In order to get a sense of how well our model would work when extrapolated to a general usage, we combined our data from all the participants to create general models for each task: reading, searching, and writing, using a 15-second time window. In addition, we also combine all data to create a universal model that is independent of the task context.

Table VI presents the results, with the accuracy measured by CCR and Kappa statistics. It can be seen that again, a multimodal model using feature fusion yields the best results, and the combination of machine- and human-specific features perform better than any of the two alone in both task-specific models and the universal model. With one exception, the models achieve Kappa statistics between 0.4 and 0.6, which gives us a good support of classifier agreement. The best classification result is 77.46% (Searching task, All features considered), 12.35% above baseline. The worst case (Reading Task, Machine-specific features only) achieves a performance of 73.07%, which is still 6.61% better than baseline.

To better understand the role of the different features, we performed a leave-one-out evaluation on all three tasks. The model was trained and tested multiple times, each time with one feature omitted.

Table VII shows the features that produced the biggest drop in accuracy when they were omitted. Both machine-and human-specific features contribute to the top 10 most contributing features. Of the mouse features, only one of them (UM – unnecessary mouse movement) makes it to the top 10. The rest are keyboard features, especially content-aware features that take linguistic information into account (e.g., letter input rate of the last word). It is also observed that most of the highest-contributing features rank highly in more than one task scenarios (highlighted in Table VII), which gives us a sense of the overall most indicative features for the general attention detection problem.

V. DISCUSSION

A. *Findings*

Our results have demonstrated the feasibility of detecting user attention and engagement from their behavior and interaction with the computer, using a nonintrusive approach and deploying only standard off-the-shelf equipment. Our models have been evaluated under different task scenarios that were chosen to be representative and realistic, and achieve significant performance gain over the baseline.

An analysis of our results shows that both fusing multiple features from multiple modalities and lengthening the time window for sampling contribute to performance improvement. Using 40 features, including 16 machine-specific features and 24 human-specific features, the results from the multimodal models almost always outperform the uni-modal models. A leave-one-out analysis shows that both machine- and human-specific features contribute to the classification result, which supports our hypothesis of the value of a multimodal approach.

In addition to feature combination, temporal-smoothing of the data also contributes to performance. This can be understood from two aspects: on one hand, lengthening the time frame produces a smoothing effect and reduces

TABLE V. PERFORMANCE OF USER-SPECIFIC, TASK-SPECIFIC MODELS

Task	**Time Interval (s)**	**Majority Baseline (%)**	**All Features (%)**	**Machine-specific Features (%)**	**Human-specific Features (%)**
Reading	*1*	63.86±6.20	67.41±2.73	59.63±7.20	66.73±3.52
	5	69.40±6.07	76.18±2.48	71.41±4.08	75.17±2.17
	10	67.04±6.30	76.44±1.97	73.23±3.12	76.55±2.22
	15	66.34±6.03	76.72±1.39	74.09±2.44	76.67±0.87
Searching	*1*	63.23±6.38	67.56±4.35	64.20±4.58	65.65±4.44
	5	69.26±8.37	76.62±2.92	74.16±2.74	75.65±2.81
	10	66.88±6.84	76.73±2.63	75.41±2.81	76.18±2.23
	15	65.91±6.76	**77.81±2.91**	76.07±2.90	76.80±2.26
Writing	*1*	66.25±5.58	70.60±3.52	69.96±3.11	67.81±3.15
	5	70.64±5.98	77.17±2.27	75.71±2.31	76.02±2.17
	10	69.05±5.86	77.49±1.40	76.14±2.19	76.63±1.52
	15	67.91±5.52	77.10±2.16	77.07±1.41	76.91±1.67

NOTE: Numbers are averaged across all users. The best performance for each model class is bolded.

 ISBN: 978-1-61208-325-4

TABLE VI. PERFORMANCE OF GENERAL AND UNIVERSAL MODELS

Task	ANA-15 (%)	ANK-15	MNA-15 (%)	MNK-15	HNA-15 (%)	HNK-15
Reading	76.57	0.49	73.07	0.39	76.41	0.49
Searching	**77.46**	0.53	75.89	0.49	76.17	0.50
Writing	76.95	0.48	76.48	0.46	76.31	0.46
Universal	75.15	0.46	73.41	0.41	74.31	0.44

NOTE: AEA-15/AEK-15: All features, Even distribution, Accuracy/Kappa Statistics, in 15 seconds; CEA-15/CEK-15: Computer features; FEA-15/FEK-15: Facial features.

fluctuation of the data. Intuitively, it seems reasonable that a user's attention at a particular point in time is not an isolated "event", but is in fact linked to many factors, including those that occurred shortly before the current moment.

B. *Specific or general*

Our work started off by constructing user-specific and task-specific models, and then generalized to a task-specific model and then a universal model. We found several challenges to creating a truly universal model. Facial features are highly distinct and it is very difficult to generalize without substantial training data. This is especially a challenge for our approach as we rely significantly on facial features. We plan to address this weakness in the future.

To a lesser extent, the variability of the tasks also hinders our efforts to create a universal model. Even though we specify the *nature* of the tasks, there is still substantial ambiguity present in the *content*. This may make it difficult to generalize our model across the same task for different users, as individual user differences, such as language ability or familiarity with the topic, will influence the result.

Even so, we made an attempt to build general models and received recognizable results. Interestingly, the model that is trained on data from all the participants achieves better performance than some of individual models. This can be interpreted into two ways: 1) we are not using enough data when constructing our user- or task-specific models, and so our models suffer from robustness, and 2) even though we collected individually distinct data, the algorithm is still able to generalize across different users and tasks to create a classification with accuracy close to that of the specific models. This result implies that creating universal models has its potential regardless of task type or participants.

C. *Lab or field study*

In order to mimic the real world as much as possible, our experiments were carried out in a lab setting, with all the chatting, phone calls and discussions that usually go along with it. The tasks are based on commonly-performed jobs and are familiar to the subjects. In other words, the subjects just need to transfer their normal tasks to the experimental computer and carry on as normal.

However, even though efforts have been made to simulate a real-world environment, there are still various factors affecting the results. To begin with, the subjects are not using their own machines (the presence of the key and mouse logger would have created serious privacy issues if installed on the subjects' personal machines), and they are also not in their "regular" environments, which may affect their engagement with the task. The need to wear the Mindset device to get the ground truth labels would also assuredly affect their performance! Therefore, we would assume a certain level of nervousness or anxiety that would perhaps not be present in a real-world setting.

In the future, we plan to conduct longer-term experiments that will allow users to work on their own machines in their familiar environments. Assuming that the privacy issues can be worked out, this allows us to observe user interactions in a more realistic context, which would create a more suitable model, which would be closer to a truly generalizable model.

VI. CONCLUSION AND FUTURE WORK

We have presented a nonintrusive, multimodal method to detect and classify user attention in a realistic setting. We obtain a performance (correctly classified rate) of 77.46% using a user-general model for a searching task, 12.34% above baseline, while our universal (user-general, task-general) model achieves a performance of 75.15%, which is 8.65% better than baseline.

Our analysis of the results brings us to the following conclusions: (1) A multimodal approach, as in a combination of machine-specific and human-specific features, achieves better performance than a uni-modal approach; (2) for attention detection, using a longer time window (e.g., 15s) yields better results than with shorter windows, suggesting that attention is a more stable attribute that does not change rapidly with time; (3) features contribute differently under different task scenarios; and 4) it is possible to build up

TABLE VII. TOP 10 MOST CONTRIBUTING FEATURES FOR EACH TASK

	1	2	3	4	5	6	7	8	9	10
Reading	NLIW	FF10	FF9	UM	BS	FF15	WS	MIDT	FF19	STD
Searching	FF5	FF11	UM	FF21	FF16	VLIW	NLIW	KBIDT	NWIS	FF23
Writing	NWIS	FF9	FF16	FF19	FF6	WS	KB	MR	NLIW	FF5

NOTE: Highlighted features rank among the top 10 for more than one task.

general and universal models with recognizable accuracy.

We believe that our approach has much potential, both in terms of providing a means through which to detect the user's attention, as well as a novel mode of human-computer interaction. We believe that the ability to detect the user's attention could be important in the development of a better interactive system which would be able to deduce users' attention state and adapt itself to enhance users' productivity.

In future work, we plan to investigate more deeply the contribution of the various features, including the use of linguistic and sequential models in the keyboard and the mouse movement features. We also plan to deploy our system as input for interactive systems and applications.

ACKNOWLEDGMENT

This work was performed at the CHILab of the Department of Computing and partially supported by Grants #G-U898 and #PolyU-523511 from the Hong Kong Polytechnic University and the Hong Kong Research Grants Council, respectively. We thank all our experiment subjects for their time and patience.

REFERENCES

[1] R. A. Calvo and S. D'Mello, "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications," IEEE Transactions On Affective Computing, vol. 1, 2010, pp. 18-37.

[2] R. W. Picard, Affective Computing. Cambridge, Mass: MIT Press, 1997.

[3] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," IEEE Transactions On Pattern Analysis and Machine Intelligence, vol. 31, 2009, pp. 39-58.

[4] P. Howard-Jones and S. Murray, "Ideational Productivity, Focus of Attention, and Context," Creativity Research Journal, vol. 15, 07/01; 2013/10, 2003, pp. 153-166.

[5] Neurosky (2013). NeuroSky - MindSet. Retrieved: Jan 2014. Available: http://www.neurosky.com/products/mindset.aspx.

[6] Emotiv (2013). Emotiv|EEG System|Electroencephalography. Retrieved: Jan 2014. Available: http://emotiv.com/.

[7] R. El Kaliouby and P. Robinson, "Real-time inference of complex mental states from facial expressions and head gestures," Conf. On Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. 2004, pp. 154-154.

[8] R. Bixler and S. D'Mello, "Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits," in Procs of the 2013 Int'l Conf on Intelligent User Interfaces, Santa Monica, CA, USA, 2013, pp. 225-234.

[9] P. Ekman and W. V. Friesen, "Detecting deception from the body or face," J. Pers. Soc. Psychol., vol. 29, 1974, pp. 288-298.

[10] C. Epp, M. Lippold, and R. L. Mandryk, "Identifying emotional states using keystroke dynamics," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 2011, pp. 715-724.

[11] P. Ekman and W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press, 1978.

[12] Qiang Ji, P. Lan, and C. Looney, "A probabilistic framework for modeling and real-time monitoring human fatigue," IEEE Transactions On Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 36, 2006, pp. 862-875.

[13] A. B. Ashraf et al., "The painful face – Pain expression recognition using active appearance models," Image Vision Comput., vol. 27, 11, 2009, pp. 1788-1796.

[14] G. C. Littlewort, M. S. Bartlett, and K. Lee, "Faces of pain: Automated measurement of spontaneousallfacial expressions of genuine and posed pain," in Procs of the 9th Int'l Conf on Multimodal Interfaces, Nagoya, Aichi, Japan, 2007, pp.15-21.

[15] O. Hamdy and Traor\'e Issa, "Homogeneous physio-behavioral visual and mouse-based biometric," ACM Trans.Comput.-Hum.Interact., v18, Aug 2011, pp. 12:1-12:30.

[16] M. Karnan, M. Akila, and N. Krishnaraj, "Biometric personal authentication using keystroke dynamics: A review," Applied Soft Computing, vol. 11, 3, 2011, pp. 1565-1573.

[17] L. M. Vizer, L. Zhou, and A. Sears, "Automated stress detection using keystroke and linguistic features: An exploratory study," International Journal of Human-Computer Studies, vol. 67, 10, 2009, pp. 870-886.

[18] G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind, and J. Movellan, "Dynamics of facial expression extracted automatically from video," in Conference On Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. 2004, pp. 80-80.

[19] D. McDuff, R. El Kaliouby, K. Kassam, and R. Picard, "Affect valence inference from facial action unit spectrograms," in 2010 IEEE Computer Society Conference On Computer Vision and Pattern Recognition Workshops (CVPRW), 2010, pp. 17-24.

[20] M. Pantic, "Machine Analysis of Facial Behaviour: Naturalistic and Dynamic Behaviour," Philosophical Transactions: Biological Sciences, vol. 364, Dec. 12, 2009, pp. 3505-3513.

[21] J. M. Saragih, S. Lucey, and J. F. Cohn, "Face alignment through subspace constrained mean-shifts," in 2009 IEEE 12th Int'l Conf On Computer Vision, 2009, pp. 1034-1041.

[22] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," in 8th IEEE Int'l Conference On Automatic Face & Gesture Recognition, 2008. FG '08. 2008, pp. 1-8.

[23] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: temporal and static modeling," Comput. Vision Image Understanding, vol. 91, 2003, pp. 160-187.

[24] Feng Zhou, F. De la Torre, and J. F. Cohn, "Unsupervised discovery of facial events," in 2010 IEEE Conference On Computer Vision and Pattern Recognition (CVPR), 2010, pp. 2574-2581.

[25] P. Viola and M. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, vol. 57, 05/01, 2004, pp. 137-154.

[26] D. Szafir and B. Mutlu, "Pay attention!: Designing adaptive agents that monitor and improve user engagement," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, Texas, USA, 2012, pp. 11-20.

[27] K. Crowley, A. Sliney, I. Pitt, and D. Murphy, "Evaluating a brain-computer interface to categorise human emotional response," in 2010 IEEE 10th Int'l Conference On Advanced Learning Technologies (ICALT), 2010, pp. 276-278.

[28] NeuroSky, "NeuroSky mindset instruction manual," 2009.

[29] L. Breiman, "Random Forests," Mach. Learning, vol. 45, 10/01, 2001 pp. 5-32.

[30] J. Cohen, "A Coefficient of Agreement for Nominal Scales," Educational and Psychological Measurement, vol. 20, pp. 37-46, April 01, 1960.

Gamification of a Project Management System

Darius Ašeriškis, Robertas Damaševičius
Software Engineering Department
Kaunas University of Technology
Kaunas, Lithuania
darius.aseriskis@ktu.lt, robertas.damasevicius@ktu.lt

***Abstract*—Gamification is the use of elements of game design and mechanics in serious contexts for enhancing the value of a service or a product for its users. We describe psychological foundations and social motives for gamification, its principles and concepts, game rules (mechanics) and elements of game as well as techniques and patterns of gamification. As a case study, we analyze gamification of a Trogon Project Management System. For evaluation of gamified interface, we propose to use WCAG 2.0 rules, adapted for evaluating color contrast of game layer interfaces, and System Usability Scale (SUS) to evaluate usability of gamification.**

Keywords—gamification; game design; game mechanics, interface; usability.

I. Introduction

Recently, gamification has emerged as a new trend in development of enterprise information and e-commerce systems. Gamification is the use of elements of game design (game rules, game techniques, gamified interfaces) in non-game contexts [1], such as marketing, employee performance and training, and innovation. The aim is to enhance the value of a service or business product beyond its face value, as well as to boost user engagement, loyalty, and satisfaction. Game elements such as badges, levels and scoreboards implemented on top of the actual business processes and combined with meaningful game rules that encourage competition between game players may help to achieve positive outcomes such as higher sales of a product, drive marketing or increase job performance.

In a recent survey by Pew Research Center, 53% of people surveyed said that, by 2020, the use of gamification will be widespread [2]. A well know study of Gartner claimed that by 2015, more than 50% of organizations that manage innovation processes will gamify those processes [3]. Over 70% of Forbes Global 2000 companies plan to use at least some elements of gamification for product marketing and customer retention [4]. While some of the expectations of the spread of gamification may be overhyped, there are several examples of successful gamification, which include *Idea Street* [5], a social collaboration platform that uses game mechanics, *Badgeville* [6], a platform that enables businesses to apply gamification across their web and mobile experiences; and *RedCritter Tracker* [7], an Agile Project Management service with badges, rewards, leaderboards, and real-time Twitter-style feeds. These gamified systems have some common aspects: an attractive graphical user interface; strong emphasis on social competition; and an engaging award system.

However, gamification still poses great challenges to software designers: 1) how to design meaningful and engaging game rules as well as integrate them with business rules, 2) how to create an attractive game interface, which integrate smoothly with a user interface of a serious system, 3) how to evaluate success of gamification both in terms of its usability (aesthetic aspect) and customer retainment (pragmatic aspect).

In this paper, we describe our experience in gamifying a Project Management System. We formally describe game rules formally, present its architecture and implementation of gamified user interface. Furthermore, we propose using quantitative (modified WCAG 2.0 (Web Content Accessibility Guidelines) [8]) and qualitative (SUS (System Usability Scale) based [9]) evaluation methods to evaluate gamified interfaces.

The structure of the remaining parts of the paper is as follows. Section II discusses the related work. Section III describes psychological foundations and social motives of gamification as well as concepts and principles, rules and elements, techniques and patterns of gamification. Section IV describes gamification of a Project Management System *Trogon*. Section V proposes evaluating gamified interfaces using Web Content Accessibility Guidelines (WCAG) 2.0 [8] guidelines and System Usability Scale (SUS) [9] as well as presents the results. Finally, Section V presents conclusions.

Several cases of application of gamification are described in the literature in the context of enterprise information systems (IS) such as a generic platform for enterprise gamification [10], implemented using service oriented and event-driven principles and best practices; authentication games [11] for improving user's behavior regarding security; and the demand dispatch system [12] with a special scoring system, leader boards and social competition aspects embedded into user interface.

II. Analysis of Gamification

A. *Psychological and social foundations*

Primary motivation for gamification is a psychology-based one, namely, to enhance user or customer motivation to do a job or to increase and retain addiction to a service or a product using a game as a tool.

Gamification can be explained by Fogg Behavior Model (FBM) [13], which claims that both, motivation to perform

 ISBN: 978-1-61208-325-4

and ability to perform, must converge at the same moment for a behavior to occur. Motivation must be supported by positive feedback from game mechanics that continuously triggers a user to perform specific actions and keeps him interested in the game.

Psychological foundation of gamification has been elaborated further by Wu [14], who analyzes why and how gamification is able to drive actions, and by Gnauk *et al.* [12], who studied extrinsic and intrinsic motivation and analyzed its relationship with external incentives and rewards.

Another motivation for gamification is social competition. Here, gamification is driven by the need to interact with other players and compare one's results. Thus, gamification requires introduction of real-time multi-user games with complex rules of game that have some similarity to social networking platforms.

B. Formal theory

Formally, gamified systems can be described using a theory of multi-games. Multi-Games is a class of games when each player can allocate its resources in varying proportions to play in a number of different environments, each representing a basic game in its own right [15]. Each player can have different sets of strategies for different basic games. The actors are permitted to play multiple games simultaneously. This multiplicity means that the actor must take interactions among relevant games and other players into account [16]. Gamified IS can be interpreted as a multi-game, i.e. a system of two games, where one game is a serious game (i.e., target IS) and another game is an entertainment game (i.e., gamification layer in target IS), where an action in the serious game leads to a reward in the entertainment game.

Following Grunvogel [17], each game G is a triple (S, M, F), where S is a set that represents the states of the different game objects, M is a monoid that represents an input of the players, and F is an action of the monoid M on set S as follows: $F : S \times M \to S$.

Then gamification can be described as a product of two games G_1 and G_2 as follows: $G_1 \times G_2 = (S_1 \times S_2, M_1 \times M_2, F_1 \times F_2)$, where G_1 is a serious (economical) game with tangible external actions and rewards, and G_2 is a non-serious game based on top of the G_1 with virtual actions and rewards.

C. Elements and Rules

According to Salen and Zimmerman [18], a game must have 1) Rules, 2) Players, 3) Struggle (artificial conflict), and 4) Goals (quantifiable outcomes). While the general goal of each game is a win, where can be multiple ways or elements of a game to reflect the player's path towards victory such as badges, which represent player achievements; leader-boards, which allow comparing one's achievements among multiple players; and levels, which reflect the growth of player skill.

Each game element can be described using the Frang [19] scheme as follows: Summary (visualization of an element with a proper description), Purpose, Ability, Motivation, Radoff's type(s) of fun (such as competition or exploration) [20], Dependencies with other game elements, and Importance.

More abstractly, game elements can be specified using a XML-based Gamification Modelling Language (GaML) [21], which provides a mechanism for precise definition of gamification concepts that is suitable for exchange on game mechanics. Finally, game rules connect game elements into a game layer. Such game rules can be modeled using a Petri Net based Machinations visual modeling notation [22].

D. Models and patterns

Several efforts exist at classifying and codifying recurring gamification practices and common techniques such as 1) Mechanics-Dynamics-Aesthetics (MDA) framework [20], a conceptual model of game elements; 2) game design atoms [23]; 3) Game design patterns [24], commonly reoccurring parts of game design; 4) game mechanics [25]; and 5) Game interface design patterns, common successful game design components and solutions such as badges, levels, or leader boards [26].

E. Architectural design

Gamification can be implemented using several architectural design methods:

1) As a service: a separate gamification system is developed, which provides elements of gamification to other systems as a service (e.g., *Mozilla Foundation OpenBadges* [27]);

2) As a module: a separate gamification module is developed that is integrated into a target system at a later stage of design (e.g., *EcoDriving* [28]);

3) As a plugin: a full implementation of gamification is developed that is later added to a target system without any additional effort (e.g., *Jira* [29]);

4) As a separate system: a gamification system and a target system are implemented separately and communicate with each other using messages (e.g., *TaskVille* [30]);

5) As an integrated system: an integrated system is developed which combines both target functionality as well as game behavior/mechanics (e.g., *RedCritter Tracker* [7]).

F. Integration with base system

According to Neeli [25], gamification of a business IS can be performed at different levels with respect to business activities: 1) at superficial level, the game mechanics are used independent of business activity of being performed, 2) at integrated level, the game mechanics are integrated into the business activity being performed, and 3) at embedded level, the business activity is designed based on game mechanics.

 ISBN: 978-1-61208-325-4

III. Gamification of a Project Management System

A. Base system

We analyze gamification of a Project Management System (PMS) Trogon (see Figure 1), as an example of a business IS (Information System).

B. Formal description

Following Bista *et al.* [31], gamification of a Project Management System is a tuple:

$$G = < J, B, R, F, P, W, T, I, D > \qquad (1)$$

Figure 1. Screenshots of Togon PMS

where J – jobs which were entered into the PMS; B – badges defined in the PMS; R – ratings based on the number of finished jobs; W – registered workers; F – trees which represent jobs in the project forest; P – worker points received; I – month or week time interval; T – time represented in 15 minute time intervals; and D – a function to determine difficulty of jobs.

The value of received points by a worker is a function

$$P(j) = \sum_{n=0}^{count(j)} \begin{pmatrix} T_r(j_n) \cdot b(j_n) \cdot y(j_n, j) - \\ \left(0.1 \cdot \left(T_r(j_n) - T_p(j_n)\right)\right) \end{pmatrix} \cdot D(j_n) \qquad (2)$$

Here, $P(j)$ is the number of points received by a worker in time interval, $b(j_n)$ is a badge received by a worker, $y(j_n, j)$ is a function that maps badges to points; $T_r(j_n)$ - time to complete the job j_n, $T_p(j_n)$ - planned time to complete the job, and $D(j_n)$ – difficulty of a job.

The game rules are as follows: (1) Every job can have a badge b and planned work time $T_p(j_n)$. (2) Every worker has real work time $T_r(j_n)$. (3) Every job has its difficulty $D(j_n)$. (4) Badge b is awarded if it is not withdrawn until the job status is „done“. Badge can be withdrawn by a project manager, if job quality is low or it took too long to finish. Quality assurance team members can remove the badge if there are many quality defects.

The player ratings are computed as follows:

1) A set of points is computed containing all employee points for considered time interval.

$$P = < p_1, p_2, \dots, p_n > \quad (3)$$

2) The set P is sorted by descending point count.

$$R = sort\_desc(P) \quad (4)$$

3) Badge board sort order is computed like this:

$$sort\_desc(count(b)) \quad (5)$$

4) Project forests are sorted by total forest size, which represents time it took to complete all jobs.

C. Principles of gamification

To gamify Trogon, a special solution was chosen combining the entire system with a gamification module. For this case study, a simple Project Management System was created with system gamification in mind. Gamification layer of Trogon PMS was encapsulated into a module. This gamification solution was chosen for several reasons: integration to an existing Project Management System is a too complex problem and can affect the quality of gamification; a full system implementation is necessary for the gamification module to be practically useful.

For gamification of a system we used the following steps: 1) Define game rules. 2) Allow players view all employee ratings. 3) Introduce badge system, which consists of several types of badges and a badge board. 4) Badge system was coupled with a level system. Every badge defines a skill and the more of the same type badges are collected the higher skill level received. 5) Special awards and bonuses are presented to most skilled employees as defined by the game rules.

D. Elements and rules of a game

The gamified Trogon PMS has a leaderboard, badge board and the project forest as main elements of gamification. Every element has its purpose. 1) The leaderboard creates competition between individual employees and allows to determine a game winner, which should be additionally awarded. 2) The badge board allows observing the skills of employees. In the badge board the employees are ordered by the total number of badges collected. Each badge (see Figure 2) represents a skill and has its own level. Progress between skill levels is displayed as a progress bar. 3) The project forest provided the element of scalability to represent the size of different projects.

Project forest (see Figure 3) is a visualization of teamwork, which has three distinct areas: unoccupied plot means unfinished tasks, and areas with trees represent finished jobs. Every tree represents a different time interval it took to finish the job, while different type and complexity of a tree (Figure 4) shows that the job required more time to complete it. This creates a forest view where, which a project manager can use to visually evaluate and compare the complexity of jobs performed as well as the skill of the employee.

Figure 2. Game badges and badge levels.

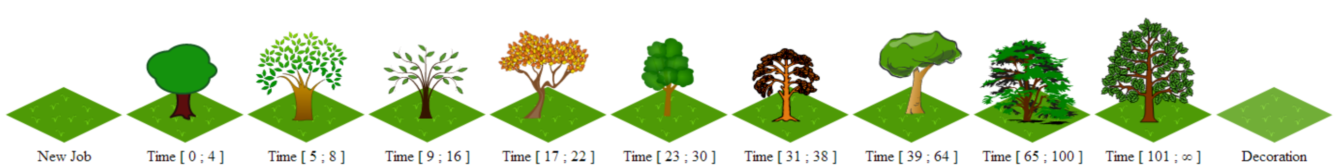


Figure 3. Elements of project forest.

Figure 4. Project forest

E. *Architecture of implementation*

The abstract architecture of Trogon PMS has three layers: 1) Website layer – this layer combines all visual elements into single system. Every website page is composed of one or more solutions from the solution layer. 2) Solution layer – contains business level logic consolidated into specific solutions. Every solution targets a specific problem. 3) Database layer – this layer is the shared by all solutions. Database maps data objects to specific tables in the relational database schema.

Gamification is one specific solution in the solution layer, and elements of gamification are used under multiple website pages. The class diagram (Figure 5) shows the division of solution into visual and logical parts. Visual and logical parts are connected by *IGamificationDC* (gamification data contract) and data object infterfaces (*IUser*, *IProject*, *IUnit* and *IBadge)*. Gamification data contract allows us to map any system, which implements the gamification data contract. In implementation: *IProject* defines all project descriptive data; *IUser* defines all user descriptive data; *IUnit* defines a unit of work, which connects project, user and badge into a single system; and *IBadge* defines all badge descriptive data. *IProjectExtensions* and *IUserExtensions* introduce computational logic to *IProject* and *IUser* data objects. Computational logic is implemented as described in formal gamification description. *BadgeBoard*, *LeaderBoard* and *ProjectBoard* are visual elements, which generate the graphical user interface for the end user to interact.

Game rules are formulated as follows. Tasks are registered and rewards for task fulfillment are assigned. Tasks are split into atomic jobs for which project manager can easily assign planned work time. Every job can hold a special skill badge. Employees enter information about their work results. Quality engineer/project manager checks completed jobs for defects, and awards badges. Employee points and badges become visible to all other employees. Every week best employee is selected to be awarded.

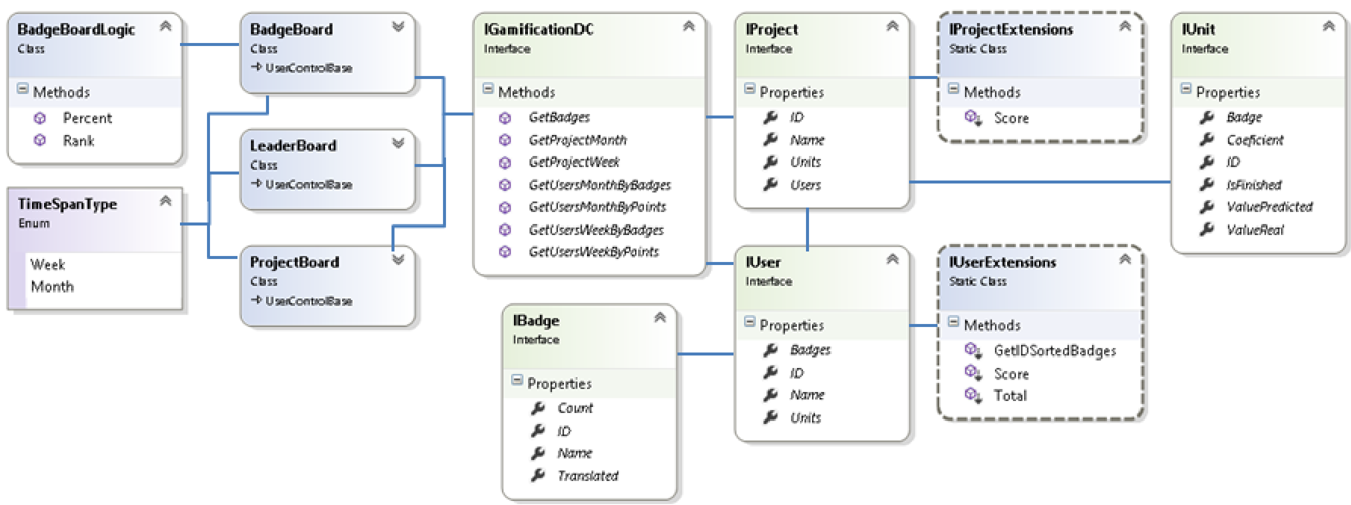


Figure 5. Gamification solution class diagram

IV. Evaluation of Gamification

Here, we describe the evaluation of gamified PMS using adaptation of WCAG 2.0 [8], and adaptation of System Usability Scale (SUS) [7] for gamified interfaces.

A. *Adapatation of WCAG 2.0 to Game Interface*

WCAG 2.0 [8] is a standard method for determining accessibility of a web interface. There are two ratings described in WCAG 2.0: the AA rating is assigned when contrast is >4.5, and AAA is assigned when contrast is >7.

Usually the WCAG 2.0 requirements are used for text only, but in our case most of information is presented in images, therefore we extend these rules on graphical images.

We use the following WCAG 2.0 evaluation scheme:

1) If the number of colors conforming to WCAG 2.0 contrast requirements is larger than the number of non-conforming colors, the interface is WCAG 2.0 compliant.

2) Else if the number of colors conforming to WCAG 2.0 contrast requirements is less than the number of non-conforming colors, but not by more than 50%, then interface has small problems, which, if resolved, would make the interface WCAG 2.0 compliant.

3) Else the interface is considered to be non-compliant with WCAG 2.0.

If interface is compliant with WCAG 2.0 then:

1) If the AAA rating colors dominate then interface is WCAG 2.0 compliant.

2) If the AA rating colors dominate then interface is WCAG 2.0 compliant.

We use the following notation to describe interface compliance:

WCAG 2.0 <X% AAA, Y% AA-, Z% AA> (6)

Here, X, Y and Z are percentage value of the AAA, AA-, and AA rating complying colors.

B. *Results of Color Analysis*

In the color analysis part of the study, we have analyzed six images of the Trogon PMS interface:

- Dashboard page, which shows all unfinished tasks, system events and inner office communications.
- Tasks page, which displays all tasks registered in the system.
- Employee task page, which displays all tasks assigned to the employee in a Gantt graph.
- Monthly ratings page, which displays the employee's ratings for the current month.
- Monthly badge page, which displays a sorted list of all employees and their badges with skill levels.
- Monthly project forest page, which displays all project forests, which had activity under this month.

We analyze screenshots (JPG images) of the game layer interfaces. For our analysis we use ImageMagick to manipulate images, Lea Verou color contrast tool to compute color contrast and define WCAG 2.0 rating, and custom script to automate the experiment.

Experiment consists of such steps: 1) We register the image of interface. 2) Using ImageMagick we generate image color histogram. 3) Using Lea Verou tool we check contrast of all colors against background color.

The tool returns one possible ratings:

- None is received when color pair is not WCAG 2.0 compatible.
- AA- is received when color pair is WCAG 2.0 AA compatible only for large elements.
- AA is received when color pair is WCAG 2.0 AA compatible.
- AAA is received when color pair is WCAG 2.0 AAA compatible.

The results of WCAG 2.0 evaluations are as follows.

1) Monthly badge board is WCAG 2.0 compliant.

WCAG 2.0 <AAA(48%), AA(23%), AA-(29%)> (7)

2) Monthly project forest is WCAG 2.0 compliant.

WCAG 2.0 <AAA(20%), AA(36%), AA-(44%)> (8)

3) Monthly leaderboard is not WCAG 2.0 compliant but with small changes compliance could be achieved.
4) Employee's task page is not WCAG 2.0.
5) Dashboard page is not WCAG 2.0.
6) Task page is WCAG 2.0 compliant.

WCAG 2.0 <AAA(69%), AA(13%), AA-(19%)> (9)

C. *Adaptation of SUS*

To rate usability of gamification we use System Usability Scale (SUS) [8] methodology. SUS already could be considered an industry standard for rating system or product usability. The main benefits of using SUS are as follows. 1) A very small number of respondents. Even with small number of respondents accurate results can be achieved. 2) Small number of questions allows a fast and efficient way to gather opinions. 3) Questionnaire can be used for system, product or module usability assessment. Drawback of using SUS is that it focuses on pragmatic quality.

Normally SUS consists of ten questions (statements), which are divided into five question (statement) pairs. In a pair both questions ask the same question, but one from positive side and the other from negative side. The SUS score is computed using such methodology: Every answer scores from 0 to 4 points. Point scale is from 1 to 5. Every question points are computed by subtracting 1 from chosen scale value. Score scale of odd questions 1, 3, 5, 7 and 9 is from 0 to 4. Score scale of even questions 2, 4, 6, 8 and 10 is from 4 to 0. The final score is obtained by multiplying score by 2.5. The total SUS score is from 0 to 100.

For evaluating usability of Trogon PMS, questionnaire consists of evaluation of game elements in Project Management System; data tables; first and second round views (ratings, badges and project forest).

The respondents are asked to respond to such statements.

1. I think what most people easily would learn game rules.
2. For me game rules looked too difficult.
3. For me gameplay elements looked easy to understand.
4. I think what I would need an experts help to fully understand gameplay elements.
5. I would like to have the possibility to always view leaderboard.
6. The leaderboard looked too complex for me.
7. I easily understand the role of badge board in this system.
8. I would need a lot of learning before I fully understand badge board role in this system.
9. I think what project forest is easy to understand.
10. I think what project forest has a lot of imprecisions.

Every pair of questions evaluates part of system gamification and whole questionnaire evaluates usability of entire system. Every pair of questions evaluated different parts of game elements: Statements 1-2 ask for usability evaluation for game rules. Statements 3-4 ask for evaluating gameplay elements. Statements 5-6 ask for evaluating leaderboards. Statements 7-8 ask for evaluating badge board. Statements 9-10 ask for evaluating project forest.

Questionnaire also asks to provide information about the respondent:

1. Your gender.
2. Your age.
3. Do you specialize in IT sector?
4. Comments.

D. *Results of SUS survey*

60 participants were asked to participate in the survey, and 30 participants have filled the questionnaire form. The main group of respondents was from 18 to 35 years. This age interval is best suited for gamification questionnaire, because their age group is considered to be the largest player group. The questionnaire included 22 men and 8 women participants. 23 of 30 participants work directly with Information Technology (IT) systems.

Every SUS question is evaluated from 0 to 10 points and every element is covered by two questions. Therefore, every gamification element can receive from 0 to 20 points. The gamification of entire system can receive from 0 to 100 points. To evaluate gamification qualitatively, we introduce the following intervals:

- 0-30 points – gamification is unusable.
- 31-50 points – gamification usability is poor.
- 51-70 points – gamification usability is average.
- 71-90 points – gamification usability is good.
- 91-100 points – gamification usability is excellent.

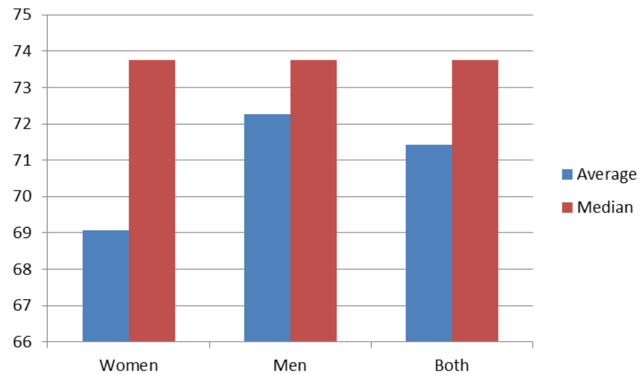


Figure 6. Gamification usability evaluation by user gender

In Figure 6, we present the results of SUS evaluation by gender. Gamification usability by genders has only small difference between women and men. The average difference is 3.5 points. We can assume what gender has almost no effect on gamification usability. Therefore, gamification of Trogon PMS is understood and evaluated pretty much without any differences between women and men.

In Figure 7, we can see large difference between evaluation gamification usability based on the experience of users working with IT systems. The difference in this case is of 17 points. The IT professionals have rated the gamification of Trogon PMS at 75 points, which is 3.9 % higher than the average rating.

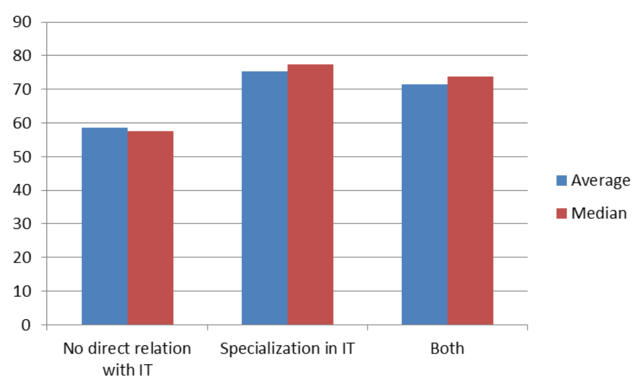


Figure 7. Gamification usability evaluation by user specialization in IT

The entire gamified Trogon PMS has been rated at 71 points from 100 points. Therefore, it has obtained “good” evaluation of gamification usability. When analyzing usability evaluations of specific elements, leaderboards were evaluated as easiest to understand, while game elements were hardest to understand.

V. Discussion and Conclusion

Gamification is a methodology that seriously enhances the systems with game mechanics and game design. In this paper, we have introduced the Trogon Project Management System (PMS) based on the three-layer architecture. Trogon PMS has been gamified by adding the additional game layer to it. The implementation of the gamification solution consists of two parts – data classes and visualization classes. This model of gamification can be applied to any system that implements the gamification data contract.

When gamifying systems there are many possible problems; so it is very important to do risk assessment as soon as possible in order to prevent against problems arising after the release of the gamified product such as user refusal. It is very important before launching gamified Project Management System into company business environment to convince employees to accept such change in companies’ policy. Trogon PMS tries to avoid such problem by introducing gamification that would not disrupt normal company workflow.

In gamified systems, the factor of motivation usually decreases over time. Furthermore, over time even the greatest games start losing attraction and pleasure they provided to their users. In Trogon PMS, we leave this problem for project manager to solve, because he is the game master. His goal is to distract from monotony and to retain motivation over time. Project manager has such tools like game prizes and gameplay elements to adjust game scenarios. For example, project manager can award best project team or organize the contest of finest forest. Short intervals of game play in Trogon PMS help to counter monotony, because long running games usually lead to motivation loss. A competition environment, which if not carefully monitored may lead to teamwork problems. This challenge is solved by not showing concrete jobs done by an employee in the team’s forest. Forest is the result of the teamwork. To avoid the problem of employees being judged by their contributions because of increased company transparency, we use information hiding to make specific job data more difficult to read.

For example, the leaderboard shows no concrete numbers but only differences between players. It allows to see only the best worker. The project forest does not actually show any concrete results for judging employees. Information is displayed only in short periods, so again you cannot judge employees. To solve a problem like tasks being not adequate, we split tasks into small concrete jobs, which are more adequate.

The experimental validation of gamification was made using quantitative and qualitative evaluation of game layer interfaces. Quantitative evaluation was made using color analysis of interfaces and its evaluation using WCAG 2.0 contrast ratio requirements for accessibility, which were

 ISBN: 978-1-61208-325-4

transferred from web page domain to game interface domain. We have proposed a new method for evaluating color contrast ratio requirements of gamification interfaces using screenshot images. The method allows to identify interfaces that are designed poorly in terms of color contrast and therefore, may not be acceptable to their users.

Qualitative assessment of gamification usability was done based on using System Usability Scale (SUS), which has been extended and adapted for evaluation of gamification usability. Our study shows that the developed system is evaluated as having good usability (71 out of 100 points). There were no usability differences between men and women. However, there were differences in evaluation of gamification usability results based on user knowledge and experience in working with Information Technologies (IT). The study has found that consumers, which were not specializing in the IT sector, have assigned a lower usability score for the Project Management System.

Future work will deal with solving the problem of integration between business logic rules and game rules, and modeling the relationship and mechanics of game elements.

ACKNOWLEDGEMENT

The work described in this paper has been carried out within the framework of the Operational Programme for the Development of Human Resources 2007-2013 of Lithuania „Strengthening of capacities of researchers and scientists" project VP1-3.1-ŠMM-08-K-01-018 „Research and development of Internet technologies and their infrastructure for smart environments of things and services" (2012-2015), funded by the European Social Fund (ESF).

REFERENCES

[1] S. Deterding, "Gamification: designing for motivation", Interactions 19, 4 July 2012, pp. 14-17.

[2] J. Anderson and L. Rainie, The Future of Gamification. Pew Research Center, May 2012.

[3] Gartner, "Gartner Says by 2015, More Than 50 Percent of Organizations That Manage Innovation Processes Will Gamify Those Processes", 2001, http://www.gartner.com/it/page.jsp?id=1629214, [retrieved: October, 2013]

[4] J. Van Grove, "Gamification: How Competition Is Reinventing Business, Marketing & Everyday Life", *Mashable*, 28 July 2011.

[5] Gartner, "DWP Case Study", http://i360institute.com/wp-content/uploads/2011/10/Gartner_DWP_Case_Study.pdf, [retrieved: February, 2014].

[6] "Badgeville", http://badgeville.com, [retrieved: February, 2014].

[7] "RedCritter Tracker", http://www.redcrittertracker.com, [retrieved: February, 2014].

[8] L. G. Reid and A. Snow-Weaver. "WCAG 2.0: a web accessibility standard for the evolving web", Proc. of the 2008 Int. cross-disciplinary conference on Web accessibility (W4A '08). ACM, New York, NY, USA, 2008, pp. 109-115.

[9] J. Brooke, "SUS-A quick and dirty usability scale," Usability evaluation in industry, 1996, pp. 189-194.

[10] P. Herzig, M. Ameling, and A. Schill, "A Generic Platform for Enterprise Gamification", Proc. of the 2012 Joint Working IEEE/IFIP Conf. on Software Architecture and European Conference on Software Architecture, 2012, pp. 219-223.

[11] C. Kroeze and M. S. Olivier, "Gamifying authentication", Proc. of Information Security for South Africa (ISSA) , 2012, pp. 1-8.

[12] B. Gnauk, L. Dannecker, and M. Hahmann, "Leveraging gamification in demand dispatch systems", Proc. of the 2012 Joint EDBT/ICDT Workshops 2012, pp. 103-110.

[13] B. J. Fogg, "A behavior model for persuasive design", Proc.of the 4th Int. Conference on Persuasive Technology (Persuasive '09). ACM, New York, NY, USA, 2009, Article 40 , 7 pages.

[14] M. Wu, "Gamification from a company of pro gamers" Lithium Lithosphere, 2011.

[15] A. Edalat, A. Ghoroghi, and G. Sakellariou, "Multi-games and a double game extension of the Prisoner's Dilemma" arXiv:1205.4973.

[16] D. L. Sallach, M. J. North, and E. Tatara, "Multigame Dynamics: Structures and Strategies," Proc. of the 11th Int. conference on Multi-agent-based simulation MABS, 2010, pp. 108-120.

[17] S. M. Grünvogel, Formal models and game design, Game Studies 5 (1) , 2005, pp. 1-9.

[18] K. Salen and E. Zimmerman, Rules of Play: Game Design Fundamentals. Cambridge, Massachusetts: MIT Press, 2004.

[19] K. Frang and R. Mellstrand, Enterprise Gamification of the Employee Development Process at an Infocom Consultancy Company. LTH, 2012.

[20] R. Hunicke, M. LeBlanc, and R. Zubek, "MDA: A Formal Approach to Game Design and Game Research", Proc. of the Challenges in Game AI Workshop, 19th National Conference on Artificial Intelligence (AAAI '04). AAAI Press, 2004.

[21] P. Herzig. Gamification Modeling Language (GaML), 15 August 2013, http://www.enterprise-gamification.com/, [retrieved: October, 2013]

[22] J. Dormans. Engineering Emergence: Applied Theory for Game Design. Universiteit van Amsterdam, 2012.

[23] B. Brathwaite and I. Schreiber. Challenges for Game Designers. Charles River Media, Boston, Ma, 2008.

[24] E. Adams and J. Dormans, Game Mechanics: Advanced Game Design, New Riders Publishing, 2012.

[25] B. K. Neeli, "A Method to Engage Employees Using Gamification in BPO Industry," Proc. of 3rd Int. Conference on Services in Emerging Markets (ICSEM), 2012, pp. 142-146.

[26] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: defining "gamification", Proc. of the 15th Int. Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '11). ACM, New York, NY, USA, 2011, pp. 9-15.

[27] "OpenBadges", http://openbadges.org, [retrieved: February, 2014].

[28] "Jira Hero plugin", https://marketplace.atlassian.com /plugins/com.madgnome.jira.plugins.jirachievements, [retrieved: February, 2014].

[29] J. N. Barkenbus, "Eco-driving: An overlooked climate change initiative.", Energy Policy, 2010, pp. 762-769.

[30] S. Nikkila, et al., "Playing in Taskville: Designing a Social Game for the Workplace.", CHI 2011 Workshop on Gamification: Using Game Design Elements in Non-game Contexts, 2011.

[31] S. K. Bista, S. Nepal, N. Colineau, and C. Paris, "Using gamification in an online community", CollaborateCom 2012, pp. 611-618.

Non Intrusive Measures for Determining the Minimum Field of View for User Search Task in 3D Virtual Environments

Zahen Malla Osman, Jérôme Dupire, Alexandre Topol, Pierre Cubaud
Centre d'Etude et de Recherche en Informatique et Communications
Conservatoire National des Arts et Métiers
Paris, France
Email: {zahen.malla_osman, dupire, topol, cubaud}@cnam.fr

***Abstract*—In this paper, we report on an experiment to determine the minimum field of view that permits the user to perform an effective search task in a 3D virtual environment, by analyzing how the user controls the virtual camera. Our study exploits a model based on the use of several novel non-intrusive temporal and quantitative measures of visual attention, such as: fixation, gaze, and movement. Seven out of ten measures gave significant results with the same findings.**

Keywords-Field of view; virtual environment; video games; visual attention.

I. INTRODUCTION

Visual attention is the ability of a vision system, whether human or artificial, to quickly select the most pertinent information from the environment in which it operates [1].

Eye tracking has been used to measure visual attention for many years. It is the process of using sensors to localize the position and the behavior of the eyes. It helps us to determine what a person is looking at, what he/she is not looking at, but also what he/she does and does not pay attention to. Through eye tracking systems we can provide many visual attention measures, such as: fixation, gaze, and movement, in order to analyze users' ocular behavior.

A principal means of interacting with 3D VEs (Virtual Environments), in the case of video games, for example, is the use of the virtual camera, which is relatively easy to access and manipulate via game engines. The use of this virtual camera can show interesting results for non-invasive study and characterization of user behavior - especially in the absence of eye tracking systems, which can sometimes be unavailable.

Our work is focused on the FOV (Field of View) effect of the virtual camera for determining the minimum FOV that allows users to perform an effective search task in a 3D VE.

The remainder of this paper is organized as follows: Section II presents related work. Section III describes our experiment that analyzes user behavior in a 3D VE via the virtual camera. Section IV summarizes our paper and provides an outlook for future work.

II. RELATED WORK

Gaming is an increasingly prevalent cultural pastime [2], and today the video game is one of the most popular types of software applications in the world. More than half of all Americans play video games, for example [3][4]. Video games can provide a framework for testing many types of attention measures [5], e.g., playing video games, such as Pac Man, can improve the reaction times of older adults [6].

During everyday interactions, our eyes provide a lot of information that reflects our emotional and mental states. Eye movement data reflect moment-to-moment cognitive processes during task execution [7]. When we look at an object in space (e.g., a wall with windows and doors), our eyes concentrate much more on some parts of this object (e.g., one of the windows), while the other parts of the object may receive less attention [8].

Studying ocular behavior in the context of human computer interaction (e.g., web browsing or video games [9][10][11][12]), allows us to identify and provide many indicators that can be used to evaluate user attention in order to improve the design of an user interface such as, for example, a digital library [13].

Much research has been conducted towards the study of ocular behavior during playing video games. In a FPS (First Person Shooter) video game, for example, the player pays more attention to the center of the screen around the reticule because he shoots enemies through the reticule; by contrast, the attention area is larger in an adventure game because the player's attention is not constrained by any specific area of the screen [14][15].

There are many types of eye behaviors: fixation, being the moment when the eyes are relatively stationary, taking in or encoding information with a minimum duration of 100 milliseconds [16]; saccade, being the eye movement that occurs between fixations with durations of approximately 150–200 milliseconds [17]; and gaze, being the moment when the eyes look at a display element [18]. When we look at an object in a visual display, we may make many fixations on this object. The number of these fixations shows the importance of the display area. A large number of fixations, however, can also reflect a poorly designed interface [18].

To study user behavior in a 3D VE, a common approach is to ask users to complete search task in order to know how he/she interacts with the 3D VE, e.g., users may have to find objects that have specified numbers displayed on them [11], or to find a maximum number of hidden keys distributed in a 3D VE [19].

Our idea was to use the virtual camera of a 3D VE to examine several visual attention measures, such as: fixation, gaze and movement. The use of the 3D VE's virtual camera provides an indirect method for analyzing the effect of FOV on user behavior, given that the useful FOV is the total area of the visual field within which individuals can obtain useful information without moving their heads or eyes [20][21].

In order to study and characterize user behavior in a 3D VE through the virtual camera, we selected several visual attention measures employed by Gibbs et al. [9]. The measures selected are expressed by the number of fixations, fixation duration, and gaze duration. We also introduced new measures to give more information about how the user performs an effective search task in our 3D VE. The measures that we added are expressed by: the number of gazes, the number of movements, the movement duration, the sum total duration of all fixations per task, the sum total duration of all gazes per task, the sum total of all movements per task, and the total duration of each task.

Our goal was to examine the FOV effect of the virtual camera on user behavior, and to determine the minimum FOV that allows the user to perform an effective search task in a 3D VE. FOV size is very important for rapid extraction and identification of information in a 3D VE. The effective search task, in the context of our experiment, consists of a simple navigation within the 3D VE for the purpose of finding all objects (e.g., hidden buttons distributed around the VE) using: the least possible number of fixations and the shortest fixation duration; the least possible number of gazes and the shortest gaze duration; the least possible number of movements and the shortest movement duration; the shortest sum total duration of all fixations per task; the shortest sum total duration of all gazes per task; the shortest sum total duration of all movements per task; and the shortest total duration of each task. Our results provide information that can be of benefit to game designers, allowing them to improve gameplay, manage the difficulty of game environments, and optimize the distribution of visual resources.

III. Our Experiment

Gibbs et al. used an eye tracking system to determine whether ocular behavior differs between newspaper websites and TV-oriented websites. They used several visual attention measures to test ocular behavior, such as: number of fixations, fixation duration, and gaze duration. Within the contest of FPS video games, our research uses these measures employed by Gibbs et al., as well as our own measures to analyze user behavior, using the VE's virtual camera instead an eye tracker. The aim of our experiment is to determine the minimum FOV that permits the user to perform an effective search task in a 3D VE, and to generate information for game designers to help them manage and adapt the difficulty of a 3D VE according to user behavior.

The users in our experiment use a mouse and a keyboard to manipulate the virtual camera of our 3D VE as they would in a FPS video game (e.g., Half Life, Counter Strike). The measures employed in our experiment, consist of various types, such as: fixation, being a short pause in movement, represented quantitatively by the Number of Fixations (NF) and temporally by the Fixation Duration (FD), which vary between 100 and 300 milliseconds; gaze, which is the time spent looking at a display object, represented by the Number of Gazes (NG) and the Gaze Duration (GD), which starts from 300 milliseconds; the movement between two fixations or gazes, represented by the Number of Movements (NM) and the Movement Duration (MD), which starts from 100 milliseconds.

We also added four measures to those specified above: the Sum Total Duration of all Fixations per task (STDF), the Sum Total Duration of all Gazes per task (STDG), the Sum Total Duration of all Movements per task (STDM), and the Total Duration of each task spent by the user to complete the required task (TD).

A total of 14 volunteers (10 male and 4 female) participated in this experiment. Their ages varied between 25 and 42 years, with a mean of 30. All participants are right-handed and healthy. The experiment was performed on a desktop personal computer with an LCD display with a resolution of 1920×1040 pixels.

A. *Procedure*

The purpose of the following experiment is to compute visual attention measures and to study the FOV effects on user behavior during a visual search task in a 3D VE. Fig. 1 shows our 3D VE, which is a virtual art gallery similar to the static environment of Lee et al. [11]. We used Unity3D version 3.5 to create our 3D VE, including all the objects and the buttons. The virtual camera is positioned at the level of the eyes of the user's avatar.

The participants were first invited to complete a short form to provide information including their name, age, gender, and whether or not they often play FPS video games. Secondly, the participants were asked to perform a free navigation in the 3D VE with a FOV of 80°, simply navigating in the 3D VE and observing the virtual objects using the mouse and the keyboard to control navigation motion. This step was created as a training phase to learn manipulation of the virtual camera. The participants used the mouse to change the orientation of the virtual camera (yaw and pitch angles) and the keyboard to move the virtual camera. We used an 'AZERTY' format keyboard with the following key mapping: Z: forward, S: back, D: right, Q: left.

Figure 1. Our 3D virtual environment (the art gallery).

Finally, the participants were asked to perform a visual search task to find and validate hidden buttons in the 3D VE. They had to find ten buttons randomly distributed on the surfaces of objects in the 3D VE (each object in our VE contains one hidden button).

Each participant had to find all the hidden buttons using the reticule area (a rectangle 250×150 pixels situated in the center of the screen), and validate them by pressing the Space key. A number is displayed at the top left of the screen to indicate how many hidden buttons are left.

The participant was asked to repeat the search task six times, knowing that we had changed the positions of all the objects and buttons, as well as the FOV size of the virtual camera before each of the six attempts at the task (10° for the first attempt, 20° for the second, 30° for the third, 50° for the fourth, 80° for the fifth, and 110° for the sixth attempt). The order of the attempts was randomized for each of the participants in order to eliminate the adaptation effect. The purpose of changing the FOV size (i.e., from 10° to 110°) was to discover how FOV affects user behavior and to determine the minimum FOV that enables the user to perform an effective search task in a FPS type 3D VE, given that the default FOV in a FPS game ranges from 75° to 110°.

B. Results

A one-way ANOVA was conducted to see whether the FOV of the virtual camera affected user behavior during the search task in our 3D VE. A total of 14 subjects took part in the experiment. We sought to discover whether there is a significant difference between the measures that we obtained by changing the FOV size between 10°, 20°, 30°, 50°, 80° and 110°. We expressed our measures by way of a natural logarithm and tested the measures' normality using the Shapiro Wilk test [22]. Then, we used the ANOVA test to analyze the variance between all our measures. We note that, for our statistical analysis, we do not take into account the random spatial distribution of objects, nor the random order of the tasks.

Table I shows the means, standard deviations and analyses of variance of all our measures. Our ANOVA results show a significant difference between certain measures used in our experiments when we changed the FOV; such as: the Number of Fixations (NF), the Number of Gazes (NG), the Number of Movements (NM), the Sum Total Duration of all Fixations (STDF), the Sum Total Duration of all Gazes (STDG), the Sum Total Duration of all Movements (STDM), and the Total Duration of each task (TD). However Fixation Duration (FD), Gaze Duration (GD), and Movement Duration (MD) don't show any significant difference.

To determine the minimum FOV that allows the user to conduct a search task within our 3D VE, we performed another ANOVA that examined all our measures between each FOV pair that we used in our experiment. The results of this ANOVA do not show a significant difference between a FOV of 10° and a FOV of 20° (p=0.0585), but they do show a significant difference between a FOV of 10° and a FOV of 30° (p=0.015*), 50° (p=0.012*), 80° (p=0.011*), and 110° (p=0.0002***). The ANOVA results also show a significant difference between a FOV of 20° and a FOV of 110° (p=0.005**), but they do not show a significant difference between a FOV of 20° and other FOVs. Finally, this ANOVA shows that there is no significant difference between FOVs of 30°, 50°, 80°, and 110°.

Impact of FOV on the TD: We observed that the TD decreases when the FOV increases (see Fig. 2).

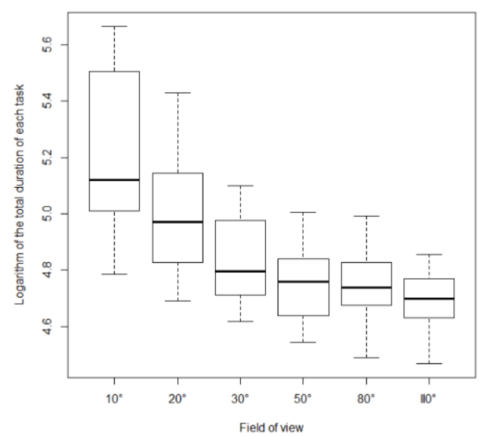


Figure 2. The Total Duration of each task (TD) by Field of View (FOV)

TABLE I. MEAN, STANDARD DEVIATIONS AND ANALYSES OF VARIANCE OF THE VISUAL ATTENTION MEASURES IN THE SIX SIZES OF FOV

	10°	**20°**	**30°**	**50°**	**80°**	**110°**	***F***	***p***
NF	2.15 (0.32)	1.84 (0.23)	1.78 (0.16)	1.66 (0.18)	1.60 (0.17)	1.57 (0.14)	14.92	<0.0001 ***
NG	1.93 (0.34)	1.66 (0.27)	1.50 (0.24)	1.45 (0.23)	1.43 (0.23)	1.40 (0.17)	8.99	<0.0001 ***
NM	2.35 (0.32)	2.06 (0.24)	1.96 (0.18)	1.87 (0.19)	1.82 (0.19)	1.79 (0.13)	12.68	<0.0001 ***
FD	2.23 (0.02)	2.23 (0.02)	2.21 (0.02)	2.22 (0.04)	2.22 (0.03)	2.21 (0.03)	1.12	0.358
GD	2.90 (0.09)	2.95 (0.12)	2.93 (0.11)	2.89 (0.10)	2.88 (0.11)	2.90 (0.12)	0.94	0.489
MD	2.44 (0.19)	2.53 (0.25)	2.49 (0.24)	2.51 (0.26)	2.57 (0.26)	2.47 (0.30)	0.41	0.838
STDF	4.37 (0.33)	*4.07 (0.21)*	3.99 (0.17)	3.89 (0.18)	3.82 (0.15)	3.78 (0.13)	15.81	<0.0001 ***
STDG	4.83 (0.40)	4.61 (0.33)	4.43 (0.32)	4.33 (0.31)	4.31 (0.32)	4.30 (0.27)	5.80	<0.0001 ***
STDM	4.79 (0.22)	4.59 (0.25)	4.45 (0.15)	4.37 (0.13)	4.39 (0.14)	4.26 (0.22)	13.91	<0.0001 ***
TD	5.20 (0.30)	4.99 (0.21)	4.85 (0.16)	4.76 (0.13)	4.75 (0.14)	4.69 (0.11)	14.84	<0.0001 ***

*** p <0.0001, ** p <0.001, * p <0.01, NF: the Number of Fixations, NG: the Number of Gazes, NM: the Number of Movements, FD: the Fixation Duration, GD: the Gaze Duration, MD: the Movement Duration, STDF: the Sum Total Duration of all Fixations, STDG: the Sum Total Duration of all Gazes, STDM: the Sum Total Duration of all Movements, and TD: the Total Duration of each task.

 ISBN: 978-1-61208-325-4

In Fig. 2, the boxplot presents the TD means of all participants in the six sizes of FOV. We found that the TD becomes convergent from a FOV of 30°. We also found that there was not much change in user behavior when he/she used a FOV of 30°, 50°, 80° or 110°; however, a FOV of 10° or 20° shows a lot of change in user behavior. For example, users took a long time to complete the task when they used a FOV of 10° or 20°, while they took less time when they used other FOVs.

Impact of FOV on the NF: We also note that the NF measure decreases when the FOV increases (see Fig. 3).

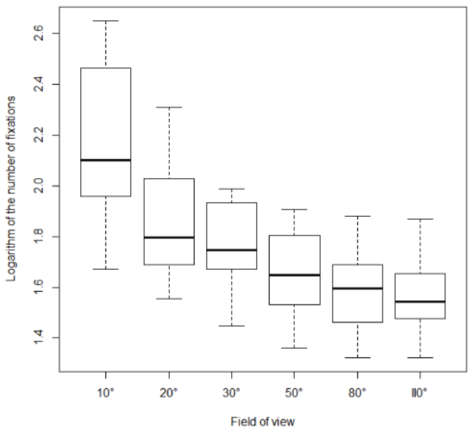


Figure 3. The Number of Fixations (NF) by Field of View (FOV).

Impact of FOV on the STDF: We also found that the STDF becomes convergent from a FOV of 30° (see Fig. 4).

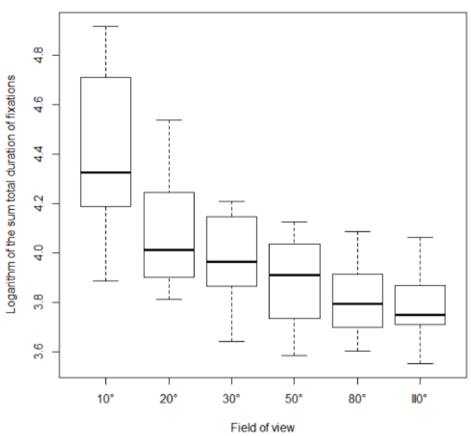


Figure 4. The Sum Total Duration of Fixations (STDF) by Field of View (FOV).

Impact of FOV on the NG: The NG in the fourth task (FOV = 50°) was high compared with the third, fifth, and sixth tasks (respectively: FOV = 30°, 80°, 110°). This is because users had difficulty in finding the hidden buttons in this task (see Fig. 5).

Impact of FOV on the STDG: We observed that the STDG in the sixth task (FOV = 110°) was high compared to the fifth task (FOV = 80°), due to the use of a large FOV (see Fig. 6).

Impact of FOV on the NM: We also found that the NM in the fourth task (FOV = 80°) was high compared with the third, fifth, and the sixth tasks (respectively: FOV = 30°, 80°, 110°) (see Fig. 7).

Impact of FOV on the STDM: We observed that there is a user in the third task (FOV = 30°) that is out the boxplot. This is because the user had difficulty in manipulating the virtual camera (see Fig. 8).

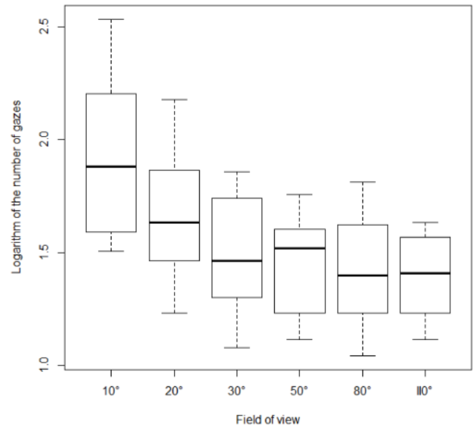


Figure 5. The Number of Gazes (NG) by Field of View (FOV).

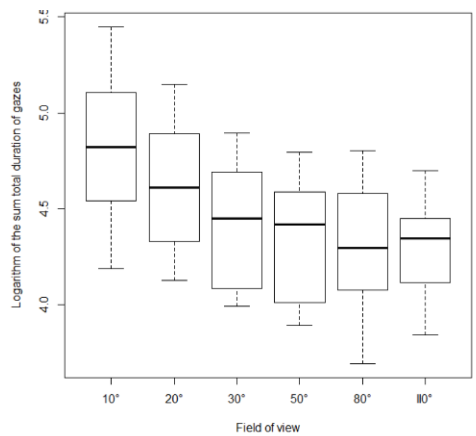


Figure 6. The Sum Total Duration of Gazes (STDG) by Field of View (FOV).

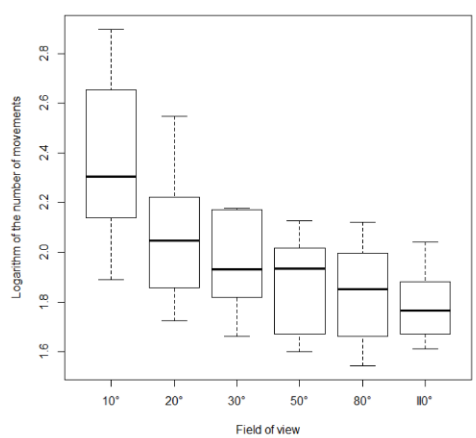


Figure 7. The Number of Movements (NM) by Field of View (FOV).

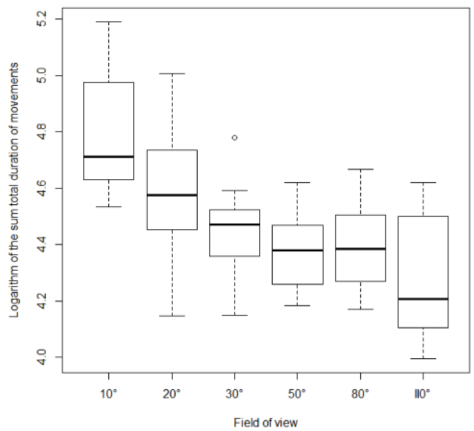


Figure 8. The Sum Total Duration of Movements (STDM) by Field of View (FOV).

After analyzing all our participants without taking into consideration their video games experience, we divided our subjects into two categories: video game players (VGP) and non-video game players (NVGP). Fig. 9 shows a comparison between the VGPs and the NVGPs using the TD measure. We found that the non-video game players took more time than the video game players to achieve the required task.

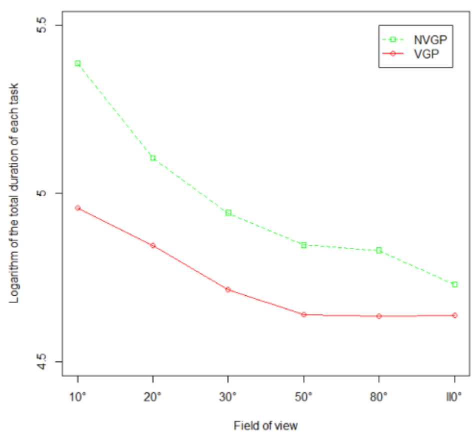


Figure 9. The difference between the gamers (VGP) and the non gamers (NVGP) using the Total Duration of each task (TD).

C. Discussion

We notice that in Table I there is a significant difference between the results based on most of our measures: the NF, NG, NM, STDF, STDG, STDM, and TD. This difference between these measures is due to the change in the FOV (i.e. between 10°, 20°, 30°, 50°, 80°, and 110°), where we observe that the FOV affects user behavior during navigation within a 3D VE. We note that these measures decrease as the FOV increases, e.g., the NF mean value for all the subjects had a natural logarithm of 2.15 when we used a FOV of 10°, and this NF decreased to 1.57 when we used a FOV of 110°. Additionally, the NG mean value for all subjects had a natural logarithm of 1.93 when we used a FOV of 10°, and this NG decreased to 1.40 when we used a FOV of 110°. We observe also that the NM mean value for all our subjects had a natural logarithm of 2.35 when we used a FOV of 10°, and this NM decreased to 1.79 when we used a FOV of 110°. We found also that the STDF decreased as the FOV increased, where the STDF mean value for all the subjects had a natural logarithm of 4.37, and this STDF decreased to 3.78 when we used a FOV of 110°. The STDG mean value for all the subjects had a natural logarithm of 4.83, and this STDG decreased to 4.30 when we used a FOV of 110°.The STDM mean value for all subjects had a natural logarithm of 4.79, and this STDM decreased to 4.26 when we used a FOV of 110°. Finally, the TD mean value for all subjects had a natural logarithm of 5.20, and this TD decreased to 4.69 when we used a FOV of 110°. The decrease is important for determining the FOV within which one can navigate effectively within a 3D VE.

The ANOVA preformed on each pair of FOVs allows us to define two groups of FOVs according to measure values: Group 1, with FOVs of 10° and 20°, and Group 2 with FOVs of 30°, 50°, 80° and 110°, given that there is not a significant difference between a FOV 10° and a FOV 20°; and between a FOV 30°, 50°, 80°, and 110°; but that there is a significant difference between a FOV of 10° and a FOV of 30°, 50°, 80°, and 110°; and between a FOV of 20° and a FOV of 110°. User behavior in Group 1 was less effective than user behavior in Group 2 because users in Group 2 performed the search task quicker than users in Group 1 with: the least possible number of fixations, the least possible number of gazes the least possible number of movements, the shortest sum total duration of all fixations per task, the shortest sum total duration of all gazes per task, the shortest sum total of all movements per task, and the shortest total duration of each task. We found that these measures become convergent from a FOV of 30°. We note that the user can use a FOV of 30° as a minimum FOV for performing the search task in a short time with minimum movement of the virtual camera. We observe also that the user can perform an effective search task using this FOV of 30° in cases where we did not find much change in user behavior based on the virtual camera when he/she uses a large FOV, such as 80° or 110°.

Finally, we see also in Fig. 9 that the NVGPs have spent more time than the VGPs to achieve a visual search task in a 3D VE, and therefore we can deduce that the VGPs perform better on the required task than the NVGPs because the VGPs are accustomed to playing video games.

IV. Conclusion and Future Work

In this paper, we have presented our experiment for determining the minimum field of view that permits the user to perform a search task in a 3D virtual environment using the virtual camera, which is accessible in all game engines.

We used several non-intrusive visual attention measures to monitor user behavior. Our results, which are based on the use of a virtual camera of a 3D virtual environment, show differences in user behavior resulting from differences in the field of view.

The participants in our experiment could perform an effective search task better when the visual attention measures' values were smaller.

We have shown that the field of view of the virtual camera affects user behavior during navigation within a 3D virtual environment to complete a visual search task. Our quantitative and temporal measures were evaluated by changing the field of view of the virtual camera. We found that the user needed less time to achieve his/her visual search task if he/she used a large field of view. We showed that the minimum field of view for performing an effective search task in a 3D virtual environment is 30°. Finally, we showed that video game players perform better in the 3D virtual environment.

Our model can be used in the context of video games to give additional information to game designers about the improvement of gameplay and management of difficulty by modifying the field of view of the virtual camera relative to difficulty level or player's needs.

For future work, we plan to prototype a First Person Shooter video game to show how our model can be of benefit to game designers. We will also test how our model can be used in the service of cognitive rehabilitation: specifically, for facilitating search tasks and adapting the difficulty of a 3D virtual environment.

Acknowledgment

This work has been partly funded by the ANR CONTINT Transread project. We would like to thank K. Neil and M. Ouattara, and all the people who accepted to test and give feedback on our experiment.

References

[1] H. Hugli, T. Jost, and N. Ouerhani, "Model Performance for Visual Attention in Real 3D Color Scenes," Proc. of the first international work-conference on the Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach (IWINAC 05), 2005, vol. 3562, pp. 469–478, doi:10.1007/11499305_48.

[2] C. S. Green and D. Bavelier, "Action video game modifies visual selective attention," Nature Publishing Group, 2003, vol. 423, pp. 534–537.

[3] A. Lenhart, S. Jones, and A. R. Macgill, "Adults and video games," Pew Internet and American Life Project, 2008, Available: www.pewinternet.org [retrieved: January, 2014].

[4] Entertainment Software Association (ESA). Essantial facts about the computer and video game industry. ESA EF 2013.

[5] W. R. Boot, A. F. Kramer, D. J. Simons, M. Fabiani, and G. Gratton, "The effects of video game playing on attention, memory, and executive control," Acta Psychologica, 2008, vol. 129, pp. 387–398.

[6] J. E. Clark, A. K. Lanphear, and C. C. Riddick, "The Effects of Videogame Playing on the Response Selection processing of Elderly Adults," Journal of Gerontology, 1987, vol. 42, pp. 82–85.

[7] K. Rayner, "Eye Movements in Reading and Information Processing: 20 Years of Research," Psychological Bulletin, 1998, vol. 124(3), pp. 372–422.

[8] A. Yarbus, "Eye movements and vision," Plenum Press, 1967.

[9] W. J. Gibbs and R. S. Bernas, "Visual Attention in Newspaper versus TV-Oriented News Websites," Journal of Usability Studies, 2009, vol. 4, pp. 147–165.

[10] J. H. Goldberg, M. J. Stimson, M. Lewenstein, N. Scottand, and A. M. Wichansky, "Eye tracking in web search tasks: design implications," Proc. of the 2002 Symp. Eye Tracking Research & Applications (ETRA 02), 2002, ACM, pp. 51–58, doi:10.1145/507072.507082.

[11] S. Lee, G. J. Kim, and S. Choi, "Real-time tracking of visually attended objects in interactive virtual environments," Proc. of the ACM Symp. Virtual Reality Software and Technology (VRST 07), 2007, ACM, pp. 29–38, doi:10.1145/1315184.1315187.

[12] Y. Kenji, W. Katsumi, and K. Takashi, "Dynamic evaluation of distribution of visual attention during playing video game," Proc. of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE 06), 2006, Article No. 96, doi:10.1145/1178823.1178934.

[13] P. Cubaud, A. Topol, and J. Dupire, "Using Game Engines for non 3D Gaming Applications," Proc. of the International Conference on Computer Games (CGAMES 05), 2005, pp. 304–307.

[14] M. S. El-Nasr and S. Yan, "Visual attention in 3D video games," Proc. of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE 06), ACM, 2006, Article No. 22, doi:101145/1178823.1178849.

[15] A. Kenny, H. Koesling, D. Delaney, S. Mcloone, and T. Ward, "A Preliminary Investigation into Eye Gaze Data in a First Person Shooter Game," Proc. 19th European Conference Modelling and Simulation (ECMS), 2005, vol. 5, pp. 146–152.

[16] P. M. Fischer, J. W. Richards, E. J. Berman, and D. M. Krugman, "Recall and Eye Tracking Study of Adolescents Viewing Tobacco Advertisements," Journal of the American Medical Association, 1989, vol. 261(1), pp. 84–89, doi: 10.1001/jama.261.1.84.

[17] L. J. Muir and I. Richardson, "Perception of sign language and its application to visual communications for deaf people," The Journal of Deaf Studies and Deaf Education, vol. 10(4), 2005, pp. 390–401, doi:10.1093/deafed/eni037.

[18] R. J. K. Jacob and K. S. Karn, "Eye tracking in Human-Computer Interaction and usability research: Ready to deliver the promises," The mind's eye: Cognitive and applied aspects of eye movement research, Elsevier, 2003, pp. 573–605.

[19] S. Hillaire, A. Lecuyer, T. Regia-Corte, R. Cozot, J. Royan, and G. Breton, "A real-time visual attention model for predicting gaze point during first-person exploration of virtual environments," Proc. ACM Symp. Virtual Reality Software and Technology (VRST 10), ACM, 2010, pp. 191–198, doi:10.1145/1889863.1889907.

[20] K. K. Ball, B. L. Beard, D. L. Roenker, R. L. Miller, and D. S. Griggs, "Age and visual search: Expanding the useful field of view," Journal of the Optical Society of America, 1988, vol. 5, pp. 2210–2220, doi:10.1364/JOSAA.5.002210.

[21] K. Murphy and A. Spencer, "Playing video games does not make for better visual attention skills," Journal of Articles in Support of the Null Hypothesis, 2009, vol. 6, pp. 1–20, Available: www.jasnh.com [retrieved: January, 2014].

[22] S. S. Shapiro and M. B. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," Biometrika, 1965, vol. 52, pp. 591–611.

Creating a Social Serious Game

An interdisciplinary experience among computer scientists and artists from UNLP Faculties

Javier F. Díaz, Laura A. Fava, Luciano J. Nomdedeu
Research Laboratory on New Information Technology - LINTI-
Informatics Faculty, La Plata National University, Buenos Aires, Argentina
jdiaz@unlp.edu.ar, lfava@info.unlp.edu.ar, luciano.nomdedeu@gmail.com

Carlos Pinto, Yanina Hualde, Leo Bolzicco, Vicente Bastos Mendes Da Silva
Multimedia Department
Fine Arts Faculty, La Plata National University
La Plata, Buenos Aires, Argentina
cpintius@hotmail.com, yaninahualde@gmail.com, leobolzicco@gmail.com, vicentebms@gmail.com

***Abstract*—This article describes the interdisciplinary work carried out by teachers and students of the Faculties of Fine Arts and Informatics in La Plata city, to develop a serious game for social networks related with Argentine native peoples. The game presented is a serious video game, innovative for social sciences, which promotes more effective learning processes than traditional methods, adding to social networks the ability to transmit knowledge, besides favoring socialization, cooperation and entertainment. The artists, computer scientists, anthropologists and primary school teachers who are part of the team believe that this game help understand and become aware of the historical and current issues of native peoples, seeking to strengthen the concept of Argentina as a multiethnic and multicultural country. The most important aspect of this serious game, called *Raices*, is to provide an innovative interactive interface to encourage the interests of school children in learning cultural history of Argentinean aboriginal communities. Also, the paper would be a guidance for other teams who would like to do similar work.**

Keywords-Serious Game; Social Game; Heritage Culture; Games for kids; Games with kids.

I. INTRODUCTION

The use of multiple electronic media is an integral part of the lives of many children today. The TV, dominant media from the '90s to the present, is having significant competitors [1][2][3]. A child today can watch a TV program on his computer, use his cell phone to browse the Internet or connect to social networks from his Tablet. While this is not the reality for all children, a large percentage has access to one or another media and therefore spends many hours using them and playing through them. Our work with heterogeneous children communities reveals that children between 9 and 12 play an average of more than 3 hours a day. Also, the number of users who join social networks like *Facebook* and *Twitter* is growing daily [4][5]. Technological convergence, a hallmark of the use of the media today, allows children and adolescents in a greater extent to access the same places through different media, often to social networks. In this context, we intended to develop a game that benefits from social networks, creating a serious game where kids have fun, share experiences with the game, compete for scores in a collaborative and collaterally, learn without trying, without seeking it. Social networks have the potential to make this happen. They help promoting a pro-social behavior, increased social skills, and provide an attractive place to play, a significant space to transmit and acquire knowledge.

In Argentina, there are nearly one thousand communities established in different parts of the country with a self-recognized native population which is 2.38% of the total national population, representing a total of 955,032 people, according to the official report of the 2010 census INDEC [6]. It is also important to stress, extending ourselves geographically, that even though in Latin America there are nearly thirty million natives, a high percentage of children are unaware of their existence or have a distorted understanding about them. It is the aim of this game to introduce the culture of these people, help so that they are perceived as people who exist today and not something of the past, helping to make their values remain, knowing that their identity is constructed and reconstructed in the context of their relationships and situations, both within the community and outside of it.

It is important to highlight that there are other games related to aboriginal communities as *Papakwaqa* [7] and *Expedition Conquistador* [8]; however, *Raices* focuses on a social approach, language, technical aspects, and is also oriented - though not exclusively- to Argentinean children and adolescents. The former is an educational game whose aim is to teach about the history and culture of Atayal tribe from Taiwan. This is an interesting game but it only relates to one community of that country and is only available in Chinese. The latter is a strategy non-educational Spanish game for children and adolescents which recreates the genocide perpetrated by conquerors in America during the conquest. *Raices* deals with local issues connected to the primary school curricula which is interesting for teachers since it could be used by students inside or outside the classroom as educational material favoring a learning process through the same multimedia they are used to.

This paper is organized as follows: after proposing a definition for serious games and discussing the current characteristics of social networks, the motivations for the development of social play are described. Then, the subgenres that make up the game and the artistic aspects are described, both visual and sound aspects. The article ends with a description of a prototype testing and conclusions.

II. Serious Games Definition

It is interesting to recover beginning the definition given by Salen and Zimmerman [9], who argue that: *"A game is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome"*. This definition has important features, such as the virtual conflict, the rules that must be met to solve the conflict and it also states that the game should have a measurable outcome, that is, that an appraisal of the possible outcomes can be done. It is also interesting to work with Juul's [10] definition based on these aspects, but which discusses in more detail the relationship between the player and the game, the game outputs are classified as positive or negative, states that an effort is done by the player to achieve something and believes that a player will have different moods depending on whether he achieves a positive or negative result. Thus, the definition given by Juul is: *"A game is a rule-based formal system with a variable and quantifiable outcome, where different outcomes are assigned different values, the player exerts effort in order to influence the outcome, the player feels attached to the outcome, and the consequences of the activity are optional and negotiable"*.

Based on this latter definition, we propose incorporating the following premise for the game to be serious: *the game should allow the acquisition of benefits in the real world*. The benefits will depend on the goal of the game, for example, educational games should enable the acquisition of knowledge, health games should promote physical and / or mental improvement of a patient, business games may favor the incorporation of labor skills for employees, etc. In particular, the game we are presenting, will try to benefit non-aboriginal children and adolescents with the acquisition of knowledge through social networks. It will also seek to raise awareness on the values of native cultures and the rights of these peoples. On the other hand, it is expected for children of native communities to be satisfied with the game and can self-recognize as such, without fear of discrimination.

III. Reasons for Creating a Serious Social Game

There are many motivations for undertaking the design and implementation of a new serious video game for *Facebook*. On one hand the high number of hours that children and teenagers spend online, sharing information with friends and playing video games is well known. This statement is based on a recent survey conducted for the National Ministry of Education which revealed that 70% of Argentine children are part of a social network and use the Internet to communicate with friends, have fun and spend time [11]. Also, to discover the habits, customs and preferences of Argentine school children between 9 and 12 years old regarding games, a quantitative methodology was applied using a survey in public and private urban educational institutions of our country. An illustrated questionnaire with 12 age-appropriate questions for scholars was designed and was provided to about 300 children in different schools by the teachers of each course. Below, the most significant data that influenced the decision to make a serious social game are shown in a graphic.

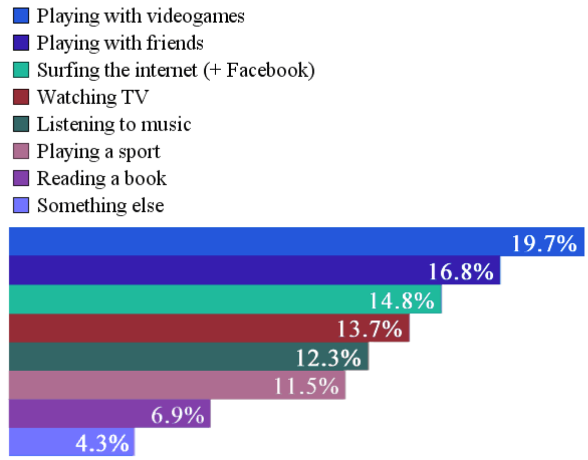


Fig. 1. Favourite activities

The graph in Fig. 1 shows the favorite activities of children outside school hours, where we can see a strong tendency to play video games, play with friends and use the internet.

On the other hand, from a more theoretical framework, it is interesting to recover Lazzaro's analysis [12] regarding the ability of games to generate emotions and how they make game an entertaining didactic instrument. She has categorized emotions into four groups, *Hard Fun* emotions provoked with the intention of overcoming obstacles and progressing, *Easy Fun* emotions evoked by visual and sound features, *Serious Fun* emotions that arise when playing ins intended to have a meaning, to be useful for something, and finally *People Fun* which is the emotion that arises from the interaction with other players, cooperation and competition between them.

In short, the expansion of information and communication technology (ICT), the everyday use of social networks in all areas and the number of daily hours children spend playing must be used / exploited for other purposes and not just for fun. The data obtained from surveys and the undoubted emotions that social games provoke, led to the design of a serious social game that aims besides being fun, to help raise awareness among children about the problems of native peoples. The viral expansion provided by social networks will promote knowledge and values that are to be transmitted to spread more easily and reach the largest population/amount of people as possible.

IV. Features of Social Games

Social networks can stimulate collaborative learning because they facilitate the formation of effective and affective groups, enable communication within groups and help to strengthen individual and collective identities.

Most games embedded in social networks obey an asynchronous communication. This is due to the fleeting nature of how people use the social network: multiple daily and short sessions. For this reason, games have had to adapt to the player´s routines and not vice versa. Asynchronous game

 ISBN: 978-1-61208-325-4

has been successful in social networks as stated by Järvinen [6], who identifies five features causing playfulness in facebook: *physicality, spontaneity, inherent sociability, narrativity* and *asynchronicity*. But now, synchronicity is also an interesting feature that could provide benefits in the interaction of players with the social game. For example, the fact that the game motivates real-time interaction reinforces the sense of "social presence" of children. Players choose social games to become more social. The immediate physical sense of the presence of other players, although in the form of avatars, is one of the factors favoring the creation of a strong sense of social immediacy. This accompanied by the immediate reciprocity that is achieved with this type of interaction, favor retention of the player in the game.

Data collected on game modalities shows a strong preference to share a common space simultaneously playing with other people (mostly friends). Moreover, it is observed that many children also like to play alone. Fig. 2 illustrates the data collected.

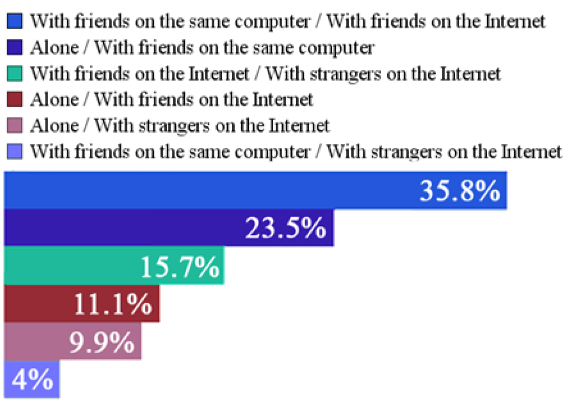


Fig. 2. Preferences regarding game modalities

Attending to the data on the polls, the game we are presenting, called *Raíces*, will contemplate different game modalities: playing alone, with friends and with strangers. To spread the word on the game, several of the functionalities provided by *Facebook* will be used, such as publishing the player's progress on their wall, showing the friends that are using the game and the level they are on, ranking the players, enabling gift sending and invitations, among others [14]. These are *asynchronous* functionalities. However, the game will also exploit some *synchronic* characteristics such as inviting players to the game to solve levels collaboratively, and communicating through virtual chats.

V. GAME GENRE

After analyzing the game genres most chosen by children at the time of having fun and considering that the video game would be embedded in the social network Facebook, we decided to make a platform game. In such games, the character handled by the player must move carefully from left to right jumping on airborne platforms, and overcome various obstacles. This genre, unlike adventure or role games, can easily be structured in short levels and does not require a continuous game, which makes it ideal for use on social networks, where access may have a fleeting nature.

In platform games, there are different aspects that may be more or less interesting for the players. In [15], three preference patterns or subgenres are identified: combat, flow and puzzle. Based on this classification and in order to make a game that appeals to any child, it can be played in three different ways (three paths). Each of these ways will have specific mechanical elements for these subgenres.

There will also be levels which combine elements of various subgenres and require the presence of other players synchronously, these players (friends or unknown) come from a different path than ours. Real-time cooperation to achieve goals, encourages the participation of players because each participant feels he is necessary to achieve a goal. Fig. 3 shows the map of the game's levels and subgenres provided.

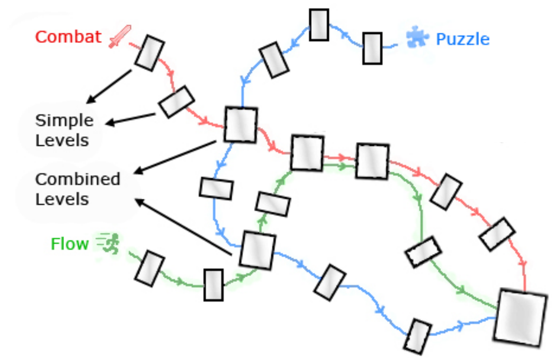


Fig. 3. Game level map

All roads have one thing in common: to complete each level certain items must be collected (level pieces) and the player must get to a door (end of level).

Below, the basic characteristics of each subgenre are described:

Combat: based on struggles against various enemies to get to the next level. These enemies are creatures that hinder the progress of the player and should be eliminated until the end of the level. The enemies in this game are characters representing people who have been hostile to the story along with the native peoples. At different levels, appear different characters in the story line, such as Spanish settlers, landowners, soldiers, harvester machinery, etc.

Flow: the player must move skillfully, jump obstacles, etc. In this subgenre the rhythmic patterns of the jumps and movements are central. Levels have more cliffs, faster movements, elements that follow the character or fall from the sky, time constraints, etc. The objective of running in the game is getting to retain aspects of native culture before a certain time, arriving before their languages are lost, arriving before their ceremonies, music, etc. are extinguished.

 ISBN: 978-1-61208-325-4

Puzzle: the player must carefully observe the level identifying which objects and paths exist, and based on this, reason as to combine together all the pieces of the level. The objective of this subgenre is favoring the incorporation of elements of different cultures by children. We will make a particular emphasis on the communion of these peoples with nature. The player must interact with various elements looking for the right strategy to solve each level.

Next, we describe the game art, visual and sound aspects that have led each subgenre to have its own aesthetics, colors and iconic characters, native music and much more, for each native people selected for this game.

VI. Game Art

Video games are a complex interplay of narrative, images and music which, from a digital platform, brings a unique experience to users. The plot provides the context in which actions take place in an environment that offers the player stimuli from an artistic discourse where different disciplines meet: music, sound, animation, painting and drawing. Below we describe sound and visual art composition of the videogame *Raíces*.

A. The sound as a narrative

According to the theorist Henri Pousseur [16], sounds must be considered as acoustic messages. They obey their own sound language syntax applied to different artistic media such as music, movies, video games, dance, etc. Henri Pousseur classifies sound messages into two types: unintentional: those that are evocative of the source producing the sound, like rain, a car passing, the wind; and intentional: those which are organized with the intention of communicating something. The latter classification can be divided into two sub types: verbal and nonverbal. Speech is part of acoustic verbal messages. Sound montage, background sound and music are considered nonverbal. Note that the background sound is the sound mimesis expected to be heard in a given area to be represented.

The sound montage is the composition of sounds that may or may not be evocative. One of the features of sound montage is the sound organization guided by the links between sounds and images. A rhythm that goes in sync with the actions of the accompanying image can be generated by sounds that evoke the represented space. The music alone, when executed by a machine or played by a musician or band, has no evocative properties, does not tell a story, just answers to the musical language and communicates through its own statement. But when being part of a sound montage it acquires contextual characteristics evoked in relation to the accompanying image: contextualizing an epoch, a feeling or a character for example.

The sound montage applied to video games serves to facilitate the player´s comprehension, accompanying his actions, capturing his attention, providing an appropriate setting to achieve the immersion of the player in a virtual world, thus reinforcing emotions. To create the sound montage of *Raíces* we opted to use sounds which evoked space and actions of the character as well as sound effects that semantically reinforce other actions without pretending to represent mimetically what is seen on the game screen.

The music of the first prototype has been made in relation to the region and culture of the native people involved in the story, with the presence of typical instruments of each region mixed with electronic effects generated by Musical Instrument Digital Interface (MIDI). The soundscape obtained is a hybrid made of two realities involved, the digital universe together with the native aspects of the country. The instruments used at this stage for the sound of the Collas uses local instruments such as *huayno, charango, the caja, quena, sikus, erkenko* and *anata.* Also we are beginning to produce the sounds for the Guarani people using the sounds of the *mimby, mbaraká, ravê, anguapú* and *takuapú* and for the Mapuche people we are using the *trutruca,* the *cultrún* the *huada* and the *kunkulkahue.*

B. The visual identity of each ethnic group

The concept art is the visual universe and product design such as a film, a comic book, a video game, etc., which involves the study, development and aesthetic coherence of the characters, fonts, sets, animations, environment and other elements. It is essential to communicate to users or viewers the genre and style, time and place, as well as the atmospheres that make the different moments of a story, the characteristics and personality of the actors or characters.

For the visual art of this videogame regarding the clothes, ethnic features, landscapes and other elements of daily life in each community, was defined a style with geometric accents. We chose an aesthetic of valued lines, outlined in black for character design, with color lines in the background and elements: clouds, mountains, trees, rocks, etc. After performing several samples and sketches, we defined geometric lines as an articulating visual motif: synthesis of strokes and shapes, typical of textile art and aboriginal pottery, which was applied to all the art elements in the game.

Once we determined the prevalent morphology and visual style in the game, a production scheme was carried out with the first sketches, where the following items were ordered according to the native group: distinctive landscape depending on the geographic location of the people, characters of both sexes, wise community elder or shaman, typical animal and finally, in a story line, the people's enemies were plotted. Fig. 4 shows this production scheme of the game.

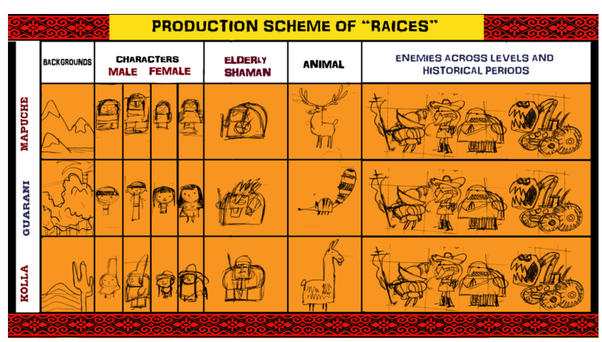


Fig. 4. Production scheme of the game

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-325-4

From this system, was determined that the backgrounds of each village would have a chromatic identity through color palettes featuring the natural landscape in which each culture settled in Argentina. They also took into account typical colors of clothes, ornaments or utensils, and a wide color symbolism in the worldview of each culture.

For the Colla people, we selected a palette of reds and oranges. This culture settled in northwest of Argentina, where hot climate prevails, so this warm range was chosen. In addition, this combination takes as reference the colors the popular area of the *Quebrada de Humahuaca* and the *Cerro de los Siete Colores*, an icon of the region, where there are predominantly reds, earth colors, yellows and oranges.

As for the Guarani people, a palette of green and yellow was selected. The chromatic combination draws on dominant colors in the jungle landscape according to the habitat of these aboriginal communities. Also these colors are present in corn, which beyond its importance as food has a strong symbolic presence. The green color was typical of the necklaces worn by women, as well as the color they chose for their headband feathers and body decorations. Wicker was an important element for building crafts and everyday objects, in yellows or golds, led us to choose these colors for the palette that identifies the natives of this region that inhabited northeastern Argentina.

Finally, for the Mapuche, a palette of blues and violets was selected. This combination adopted to identify this native people of Patagonia, was chosen taking as reference the blue seas and lakes that fully visually identify the landscape of southern Argentina and Chile where these communities lived and live. The monumental mountains and vast spaces between them generate an atmospheric perspective that makes blue colors predominate. The continental ice and snow also bring colors in these ranges. Blue is not just a visual reference, but also spiritual, blue color is associated with energy and great forces that occur in the location of the Wenu Mapu (above land).

In addition, for the design of the characters, features were synthesized considering physical characteristics and taking into account the costumes of the native peoples, giving the user the possibility to choose from a repertoire of traditional clothing of each region to customize clothing and accessories of his avatar. Fig. 5 shows the characters of the Colla people in the video game. The color palette for clothing was defined taking opposite colors from the chromatic circle to the predominant palette in the landscape to achieve color contrast and visibility.

Fig. 5. Characters and clothing of the Colla people

So, landscapes and elements of the different levels of *Raíces* refer to aboriginal weaving and textile forms that allow linking to modular geometric aesthetic and which refer to pixel art, as seen in the elderly, in Fig. 6.

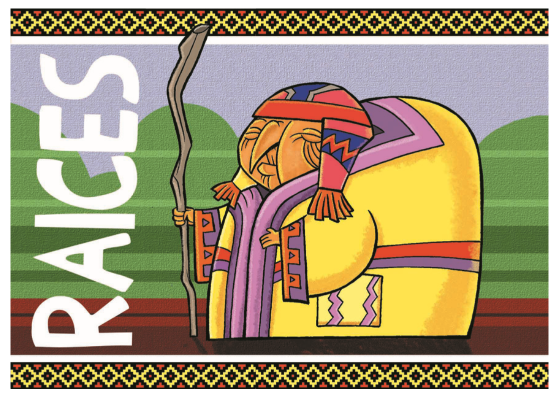


Fig. 6. Pixelated image of elderly Colla (Shamán)

This choice is based on the translation of data on daily life according to the dialogue between the cultural layer and informatics layer. Formal synthesis operations, geometrization, chromatic organization and texturized surfaces in the design of characters, objects and interactive backgrounds was defined in relation to the processes of aesthetic research and technical resolution in accordance with the codes of the gaming device .

VII. TRANSMISISION OF KNOWLEDGE THROUGH PLAY

It is intended that as they progress in the game, children can recognize and incorporate aspects of the culture of these native peoples such as language, dress, music, customs, etc. The avatar customization contributes in this direction by including traditional clothing and accessories specific to each culture, allowing children to play with different combinations and, almost without seeking it, through to descriptions, discover why and on what occasions it was used. Also each character will transit through their own geographical areas, enabling the transmission of information about the places of origin of these people.

 ISBN: 978-1-61208-325-4

Cultural aspects such as musical instruments, traditional medicine, artistic events (pottery, silverware, etc.) and religious symbols are represented in the game. The pieces to obtain at each level will correspond to these elements, allowing the player to discover through the levels the various elements that are part of each culture. For instance, when the player gets a piece representing a musical instrument, the real sound of the instrument will be reproduced, thus letting the children recognize their sonority.

Also important is the existence of cross-cutting themes that address common to all subjects. Respect for nature and the wisdom of the elderly are some of them. To reflect the first issue, respect for animals and plants will be stressed. If during the course of the game, a player causes unnecessary damage to nature, they will be warned and punished accordingly. With regard to the wisdom of the elderly Shamans, at each level there will be a character (elder) who gives advice to the player on how to move on, provides information on the meaning of the different elements and at certain levels narrates little stories.

As for language, an approach to the various sounds will be sought. Many of the items that appear will have their names in the original language as well as some words from the story of the elderly. The virtual chat necessary to communicate with other players include prefixed messages with bilingual text for children to recognize basic words.

During loading levels, time will be used to include information as a trivia offering the possibility to add additional points.

Finally, and to complete the educational aspect of the game, as the different pieces or objects are obtained, or certain stories told by the elderly are "heard", certain items will be unlocked in a kind of game album. This section can be accessed at any time to view all items achieved and the information about.

VIII. TECHNICAL DETAILS

The selection of technologies for the implementation of the videogame *Raíces* (which, as anticipated, was thought for *Facebook*) was based on two requirements: the game should work in most web browsers and the web technologies used should be open. For these reasons, we chose to develop the game in javascript and HTML5. Flash technology, widely used for this kind of development, was discarded.

After analyzing the features provided by several existing game engines, small prototypes were implemented, initially with Construct2 [18], then with Impact [19], and finally, with Turbulenz [20]. This latest open-source 2D/3D engine provides a set of functions for handling graphics, sound, user input and resources, and has a complete and efficient 2D physics engine, which encouraged its selection for the implementation of *Raíces*.

WebGL [21] is an immediate mode 3D rendering API designed as a rendering context for the HTML Canvas element. The HTML Canvas provides a destination for programmatic rendering in web pages, and allows a better performance for that rendering using different rendering APIs.

Although the use of WebGL enables superior performance, not all browsers support this technology and Turblenz graphic renderer works exclusively on WebGL. To solve this problem and in order to look for greater compatibility, the graphics engine management was replaced by the Pixi.js [22] library, a 2D canvas with WebGL renderer fallback. Thus, Pixi.js automatically identifies the best option: the rendering is done through WebGL when this technology is available and through the Canvas object for the case of non WebGL browsers.

Turbulenz engine also provides facilities to manage leaderboards, badges, networking and multiplayer. Although these features are heavily tied to Turbulenz game platform, they will not be used. Instead, the server-side part of the game will be supported through NodeJS, Javascript-based platform for data-intensive applications Real-time [23].

The integration of *Raíces* with Facebook will be done through the Javascript Software Development Kit (SDK) provided by the platform [24]. This SDK, among other things, allows Facebook Login and manage calls to the Graph API interface from which the application can read and write data and interact with the different components of the social network.

IX. TESTING OF A FIRST PROTOTYPE

To get a first impression of the impact caused by the basic game mechanics, the dynamics generated from them and the emotional responses of children when they interact with the game, children between 9 and 13 years were convened at the Faculty of Informatics for testing a first prototype implemented with the three genres.

Fig. 7 shows video scenes filmed during the game testing session, where it is possible to observe that the game provokes emotions and a high concentration.

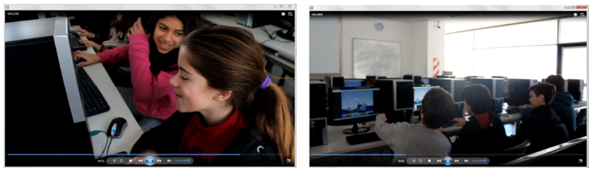


Fig. 7. Prototype testing

There was enthusiasm in passing to the next level and reaching the end of the game. Some faces of frustration were noted in gender *thinking* and less excitement in the *run*. This will be taken into account as feedback on the new stages of development. During the testing they were given a short survey to determine the feelings caused by the different genres. Fig. 8 shows some results. Also as part of the survey, 33% responded that the puzzle was the subgenre they liked most, 45% chose combat and 18% chose the subgenre run. While there are differences between them, each subgenre has its adherents.

 ISBN: 978-1-61208-325-4

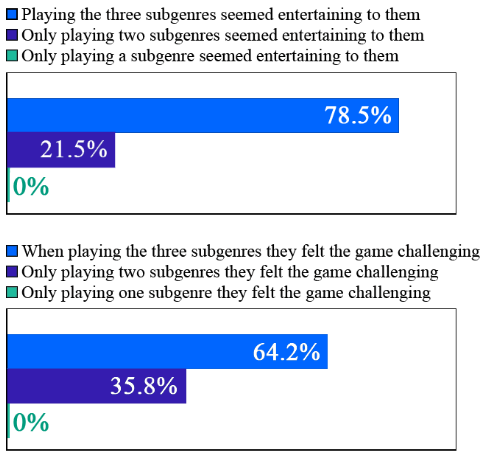


Fig. 8. Emotions caused by the subgenre

After the testing, the children asked to continue playing and were interested in the new levels and stories. Following this, a group on Facebook was created that informs about new versions of the game and the kids comment and suggest new features. This is a way to incorporate them into the game design. In this line of work where a game for children and with children is designed we are also working with a plastic workshop of our city [25], where teenagers have drawn scenes from legends of native peoples with the topics covered in the video game, which then will be collated in legend and may be accessed on demand from the context of the game.

So far, two evaluations of prototypes have been carried out where the game mechanics and the emotions generated by the game were analyzed. In later versions of the game, we will also evaluate if the knowledge the game is trying to convey is acquired by the students, that is, if learning is favored by playing with *Raíces*. For this reason, we are working very closely with teachers from different schools participating in the project, so that we can compare if the learning process in students playing with *Raíces*, is different from that of students who are not.

X. CONCLUSIONS

After long struggles and silences, the aborigines of Argentina are starting to recover their place and rights as native peoples. However, many problems persist and the contribution of education to a final recognition of their rights and full integration is essential. This paper has presented the design of a serious game for *Facebook* that provides an attractive space to promote socialization, cooperation and entertainment for children, while collaborating with the acquisition of knowledge. It is also expected to help raise awareness of the historical and current problems of native peoples.

Regarding the design of the game, we have described the theoretical frameworks taken into account to manage to make it fun and encourage children to play with it, we have also described the game art, both visual and sound aspects.

Finally, it should be pointed out that the data collected from surveys of our own and others confirm that children and young people today are very crossed by audiovisual culture and consumption associated with new technologies making it positive to leverage these characteristics / traits / peculiarities when generating new educational proposals.

It is expected that this videogame, mark of the time, will be conducive to more effective learning processes than traditional teaching methods, and that along with teachers from educational institutions we may demonstrate the effectiveness of this innovative approach.

XI. ACKNOWLEDGMENTS

We are grateful to the educational institutions and teachers that made it possible to conduct the surveys.

REFERENCES

[1] M. Castells, "The Culture of Real Virtuality: The Integration of Electronic Communication, the End of the Mass Audience, and the Rise of Interactive Networks", in The Rise of the Network Society: With a New Preface, Volume I, Second edition With a new preface, Wiley-Blackwell, 2010, Oxford, UK. doi: 10.1002/9781444319514.ch5

[2] "Digital-Set-Surpass-TV-Time-Spent-with-US-Media", available at: http://www.emarketer.com/Article/Digital-Set-Surpass-TV-Time-Spent-with-US-Media/1010096, [Last accessed: February 20th, 2014]

[3] E. Respighi, "Baja el rating, sube la web", available at: http://www.pagina12.com.ar/diario/suplementos/espectaculos/8-31237-2014-02-04.html. [Last accessed: February 20th, 2014]

[4] J. Constine, "Facebook's Q2: Monthly Users Up 21% YOY To 1.15B, Dailies Up 27% To 699M, Mobile Monthlies Up 51% To 819M", TechCrunch, available at: http://techcrunch.com/2013/07/24/facebook-growth-2/. [Last accessed: February 20th, 2014]

[5] M. Watanabe and T. Zuzumura, "How social network is evolving?: a preliminary study on billion-scale twitter network", International World Wide Web Conferences Steering Committee Republic and Canton of Geneva, Switzerland. table of contents ISBN: 978-1-4503-2038-2, pages 531-534.

[6] Official 2010 Census Report of the National Institute of Statistics and Census of Argentina (Spanish: Instituto Nacional de Estadística y Censos, INDEC), updated on 2012. Available at: http://www.censo2010.indec.gov.ar/cuadrosDefinitivos/analisis_cuarta_publicacion.pdf. [Last accessed: February 20th, 2014]

[7] H. Chih-Hong and H. Yi-Ting, "An Annales School-Based Serious Game Creation Framework for Taiwanese Indigenous Cultural Heritage", Journal on Computing and Cultural Heritage (JOCCH) - Special issue on serious games for cultural heritage JOCCH. Volume 6 Issue 2, May 2013, Article No. 9.

[8] Expeditions/Conquistador, available at: http://conquistadorthegame.com/. [Last accessed: February 20th, 2014]

[9] K. Salen, and E. Zimmerman, "Rules of Play: Game Design Fundamentals". 2004, The MIT Press. ISBN=13-978-0-262-24045-1.

[10] J. Juul: "The Game, the Player, the World: Looking for a Heart of Gameness". Level Up: Digital Games Research Conference Proceedings, edited by Marinka Copier and Joost Raessens, 30-45. Utrecht: Utrecht University, 2003.

[11] M. Carbajal. "The generation of the multiple connections", available at: http://www.pagina12.com.ar/diario/elpais/1-199322-2012-07-23.html. [Last accessed: February 20th, 2014]

[12] N. Lazzaro, "Why we Play Games: Four Keys to More Emotion in Player Experiences", proceedings of GDC 2005.

[13] A. Järvinen, "Game design for social networks: interaction design for playful dispositions". Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games, New Orleans, Louisiana., 2009.

[14] X. Wei, J. Yang, L. A. Adamic "Diffusion dynamics of games on online social networks", WOSN'10 Proceedings of the 3rd conference on Online social networks, 2010.

 ISBN: 978-1-61208-325-4

[15] N. Nygren, J. Denzinger, B. Stephenson, J. Aycock, "User-preference-based automated level generation for platform games". 2011 IEEE Conference on Computational Intelligence and Games (CIG).

[16] H. Pousseur, "Semantic music and society", Alianza Editorial, Madrid, 1984.

[17] M. Carabajal, "La generación de las conexiones múltiples" available at: http://www.pagina12.com.ar/diario/elpais/1-199322-2012-07-23.html. [Last accessed: February 20th, 2014]

[18] Construct2 Game Engine: https://www.scirra.com/construct2. [Last accessed: February 20th, 2014]

[19] Impact Game Engine: http://impactjs.com/. [Last accessed: February 20th, 2014]

[20] Turbulenz Game Engine: http://biz.turbulenz.com/developers. [Last accessed: February 20th, 2014]

[21] WebGL, WebGL Working Group: http://www.khronos.org/webgl/. [Last accessed: February 20th, 2014]

[22] Pixi.js Renderering Engine, available at: http://www.pixijs.com/. [Last accessed: February 20th, 2014]

[23] Node.js, available at: http://nodejs.org/. [Last accessed: February 20th, 2014]

[24] Facebook SDK for JavaScript, available at: https://developers.facebook.com/docs/javascript. [Last accessed: February 20th, 2014]

[25] Art Workshop "The cow with many colors", http://lavacademuchoscolores.blogspot.com, https://es-es.facebook.com/LaVacaDeMuchosColores. [Last accessed: February 20th, 2014]

What games do

Interaction, Design, and Actor Network Theory

Jari Due Jessen
Center for Playware
Technical University of Denmark
2800 Kgs. Lyngby, Denmark
jdje@elektro.dtu.dk

Carsten Jessen
Centre for Teaching Development and Digital Media
Aarhus University
2400 Copenhagen, Denmark
cj@dpu.dk

***Abstract*—When interacting with computer games, users are forced to follow the rules of the game in return of the excitement, joy, fun, or other pursued experiences. In this paper, we investigate how games achieve these experiences in the perspective of Actor Network Theory (ANT). Based on a qualitative study we conclude that both board games and computer games are actors that produce experiences by exercising power over the user's abilities, for example their cognitive functions. Games are designed to take advantage of the characteristics of the human players.**

Keywords: computer games; board games; Actor Network Theory; interaction; game research; game design

I. Introduction

Using computer software usually means the user is the active part, which controls the interaction by input and direct manipulation [1]. Interaction with computer games is a different experience, because the user acts in a game world, where the game content has excessive influence on the gamer's behavior. Game figures and other game items are not just passive objects, they can be manipulated as the gamer pleases. If a game is to come alive, the gamers have to follow rules and act as the game indicates. Playing a computer game like *Counter Strike* [2] or *World of Warcraft* [3] is not just a question of manipulating an avatar. The game is forcing the gamer to react to events in the game by acting in a certain way, if the gamer wants to survive and prosper in the game, i.e., the gamer is placed in a role he or she has to fulfill. In other words: games do something to and with people who play them, and, in a certain way, games are just like actors who have an agency. What this agency consists of and how it is engineered is of interest for designers.

In this paper, we will show how games can be seen as actors and as organizers of actors and actions on the basis of Actor Network Theory (abbreviated to "ANT") [4]. ANT is well suited for the analysis of user interaction with games because ANT offers an approach to agency that does not assign power only to human actors, but allows the possibility for objects and rules to be studied as actors. Also, ANT opens a way of seeing design as a social enterprise. As Yaneva stresses: "…design has a social goal and mobilizes social means to achieve it" [5].

ANT has received some attention in game studies during the last decade. Several scholars have studied games on the basis of ANT [6], especially focusing on the interchange between humans and technology [7] or on the development of social networks in online games [8]. In this paper, we will take a different approach and show how the ANT perspective can explain which forces are at work, when games are actually played. Thus, our focus is on what the immediate effects of using games are.

The paper is the result of a research project where we studied gamers in different ages playing computer games as well as board games. Our point of departure was that computer games are games before they are anything else [9]. Therefore the study focused on studying games as a genre rather than just digital games, and our main example in this paper is a board game.

In the next section we will introduce the ANT focusing mainly on the concept of "translation", which is employed as our main analytical foundation. After this, the paper will present the used research methodology for collecting data. In the following sections, the selected case of game playing will be presented, followed by a presentation and a discussion of the results of our investigation.

II. Actor Network Theory

ANT was first developed by science and technology study scholars Michael Callon and Bruno Latour [10] as a new approach to social theory. ANT is of interest to any analysis of technology, which goes beyond the assumption that technology is a mere instrument that we, as humans, utilize. ANT holds that any element of the material and social world (nature, technology, and social rules) can be an actor in the same way humans are. Agency is never only human or social, but always a combination of human, social and technology [11]-[13].

ANT is not a theory in the usual sense of the word, according to Latour himself, since ANT does not explain "why" a network takes a certain form or "how" this happens [4]. ANT is more a method of how to explore and describe relations in a pragmatic manner, a "how-to book" as Latour calls it [4], and thereby offers a way to describe ties and forces within a network.

The main idea in ANT is that actions always take place in interaction between actors in networks, where the actors influence each other and struggle for power. We usually see social interaction between humans this way, but ANT differs

 ISBN: 978-1-61208-325-4

from traditional social theory by stating that the actors are not only humans but can be other elements as well.

A. *The traffic example*

ANT can be hard to grasp, and even counter-intuitive [11], because it reverses our common understanding of actors and agency, for instance when it cuts across the subject-object division underlying our thought about the world, we live in. In an attempt to clarify ANT, Hanseth and Monteiro [14] use traffic as an example to explain the implications of seeing something in the perspective of ANT. We find their example very usable to give a better understanding of ANT and, hopefully, what we later have to say about what games do. The following is a short presentation of their attempt and afterwards we will use it to explain the process of translation: When you are driving in your car from one place to another, you are acting, but your acts are heavily influenced by technology and the material world (the car's maneuvering abilities, the layout of the roads, traffic signs, traffic regulation), the immaterial (traffic rules, traffic culture) and habits (your own experience as a driver) [14].

According to ANT, these factors (including you) all function as actors and should be understood as forces of agency in a linked network. Human and non-human, technical and non-technical elements are part of the network, and none of the elements are per definition granted special power over the others [11], [14].

Expanding the thoughts of Hanseth and Monteiro, we can add that, in the traffic example, you want to move from place to place, but you are dependent on technology and forced to act in accordance with both social rules and physical conditions. Even though you are the driver, you will clearly feel the forces of other actors when acting out the driving. For instance, the road forces you to follow a certain route, the traffic light forces you to stop and start. One can say that in order to reach your goal safe and fast, you have to "give in" to the network and in a way "hand over" your acting power and control over the car, so that the vehicle will move in accordance with the demands of traffic network. You have to "delegate" [11] power to the traffic network, and, in return, you will reach your goal fast and safely. Of course you are not handing over the control of yourself to the network. To delegate is more to act as prescribed by other actors. According to ANT, this is what happens in an actor-network relation.

B. *Translation*

The way delegation is done is through the process of *translation*. This process requires the actors in a network to accept roles, a worldview, rules of acting, a path to follow etc. Michel Callon [15] describes the process of translation as a process of "persuading" with four distinct phases, he calls "moments": problematization, interessement, enrolment, and mobilization. These moments are inter-related overlapping steps that describe how stable actor-networks come to be established [16]. We will introduce them briefly in the following, and later use them in our game analysis.

The first moment, problematization, is where some of the actors in the network in question bring forth a definition of the problem and present a viable solution to it for the other actors. This is also the process in which the actors' roles are defined (both human and non-human actors). To use the traffic example above, this is where the car and the traffic network are presented as a solution to the transport problem.

As part of the problematization process, a so-called obligatory passage point (OPP) is defined, i.e., a practicable solution, which the actors have to accept to achieve their goal. An OPP "is viewed as the solution to a problem in terms of the resources available to the actant [actor] that proposes it as the OPP (…) It controls the resources needed to achieve the actant's outcome." [17] By defining an OPP, other possibilities are closed [15]. In the traffic example, the OPP is literally a passage, since it's the roads and the current traffic rules etc., which have been established as a solid, reliable network.

The second moment, interessement, is where the main objective is to convince all the involved actors that the proposed problem and solution is the correct one so that they will accept to use this solution and not another one. In the traffic network, this is done by the use of sanctions from traffic rules, signs, and, not the least, by the learning processes human actors go through to get a driver's license.

When the interessement of the actors is successful, the third moment, enrollment, is happening. This moment is important since it is here that support and allies are created, and the process by which actors become part of a network. The process can happen in many ways: "To describe enrollment is […] to describe the multilateral negotiations, trials of strength and tricks that accompany the interessements and enable them to succeed." [15]. In relation to the traffic network, one can think of all the things that support cars and their moving along the roads.

Finally, the last moment, mobilization, is where the actors are mobilizes in such a way that they act in accordance with the prescribed roles and thereby they maintain the established network. This happens when the drivers drive their cars following the rules and pathways of the traffic network.

C. *Design as inscription*

The effect of translation is delegation of power and agency. I relation to design of objects, e.g. computer games, translation is about how to construct an object in such a way that users are convinced to delegate agency. This is described as *inscription* and *description* by Madeleine Akrich [18].

Inscription is the process where a designer embeds a special way the user has to interact with the designed object. The designer is envisaging a user and a use case for the object and develops an intended use, which is inscribed into the object by use of, for instance, physical shape, GUI, behavior of objects, and affordances in general.

Akrich compares inscription with a movie script, and calls the result a script for how the user should use the object. We see this, for example, in the design of the iPad's user interface, where users are compelled to use finger movements to interact, which is a more intuitive way of interacting and quite different from using a computer mouse.

While inscription is the designer's idea and framing of the interaction, Akrich uses the term description to describe the actual usage of the objects. This is where the script build into and drawn upon in the design process, meets the user in an actual user setting. Coming alive is the central part of description. It is central to ANT that a non-human actor can have agency and perform actions, and this is what we see when the scripts embedded in designed objects comes to live and the objects engage in a network with other actors.

In the perspective of ANT, a game can by studied as a designed object with inscriptions that has agency and does something with the user, because the user invokes a network of actors and agency when he or she starts playing a game, i.e., following the rules of the "game world". A game designer has to be aware of the network of actors the specific game design can invoke if he or she wants to be able to use it in the process of inscription. Networks of actors are the unit of analysis in our study presented below.

III. Research Methodology

Our research method relied on qualitative data collected through observation, both non-participatory observation and active participation [19], [20]. We collected data from 12 game sessions where we observed informants, recorded their behavior and interviewed them before, during and after playing. To ensure recordable data, we used games, where players had to be social and communicate with one another and board games was especially well suited for this, since people tend to talk more when playing such games. We observed children as well as grownups and mixed age groups playing games in natural settings at home, in the family, or with friends.

The purpose of our study was to investigate and describe agency and actors at work when gamers play games. As our framework of analysis, we employed the concept of actors and agency and the four described moments of translation, being careful not to differentiate between non-human and human actors. We recorded spoken language as well as body language and gestures, and managed data using theoretical coding as described by Uwe Flick [21]. We analyzed agency by following what people did with games, extracting actors and ties, and described the translation process in the actual game situations, as we will demonstrate in the next two sections.

IV. Case: the game "Quackle"

The case of playing the board game "Quackle" in a mixed age group is exemplary for our observations in general and in the following we will use our analysis of this case to present our interpretation of what the game actually does.

A. *Quackle! The game*

The game, which was awarded "Game of the Year" in Denmark in 2006, is a typical funny board game for the ages 5 and up. In short, the game consists of 12 different animal figures, 8 barns and 97 playing cards with pictures of the animals and one arrow card (see Figure 1). The game starts with each player pulling an animal figure from a clothed bag, showing it to the others and then hiding it in his barn so the

Figure 1. Photo of the game Quackle! with animals, cards, and barns on the left.

others can no longer see it. The cards are dealt and placed in a pile in front of each player face down.

The objective of the game is to get rid of all the cards you have in your pile. Each round of the game consists of the players in turn turning a card and placing it for all to se. If two players have the same animal on their card they enter a *battle*, where the players compete on being the first to loudly say the sound of the *other* player's animal hidden in the barn. The player that looses the battle needs to pick his own and the others pile of upwards facing cards. The game continues until there again are two identical animals in the cards, or one of the players gets rid of all their cards [22].

The game seems pretty simple, but requires that the players can remember and quickly mobilize the correct sounds when two identical cards are turned, which is more difficult than one might think, even for adults.

B. *Game inscription*

As we see in the above description of the game, there is a special way, players are expected to interact with the game (the inscription), and, as we will argue in the following, by this the game uses the learned scripts that the player brings along and the player's physical and psychological abilities. Among other things, the game takes advantage of the players' knowledge (i.e., scripts) about animals and animal sounds, and the game utilizes the fact that in pressured situations most humans have a tendency to react automatically. It is precisely this automatic reaction that makes the game fun, because the players' makes a lot of mistakes trying to be the fastest, which often result in weird sounds that is a mix between different animal sounds.

The game designer has created an inscription that can be indicated as follows: We must say a particular animal sound, while we see and try to remember a lot of other animals. These many inputs are combined with the stress factor that the game introduces by stating we must respond faster than our opponents! Thus, the inscription creates a special way the player has to act, i.e., a way the players have to use their abilities.

 ISBN: 978-1-61208-325-4

In the perspective of agency it is noteworthy that the game forces the player to make mistakes and thereby produce a mishmash of sounds, which the player would not normally produce. When we asked our informants about the experience most of them said their tongue was "out of control". In that sense, it is apparent, that the game has agency and does something to the player.

C. Translation

The inscription plays an important role when considering the whole situation as a translation. As previously described, the translation consists of four moments that we will now outline in relation to the game scenario.

The first moment is the problematization, which is where we are presented with a problem. In our case, the game is played in natural situations on a Friday evening in a family of four (parents and two children, son age 12 and daughter 21). For the family, the problem is the need for entertainment understood as a peaceful and enjoyable social time together. In this case, the game of Quackle is set up as a solution. Like any family game and most entertainment products, it promises that playing the game will lead to the experience of fun. Thus, the game is put forward as an actor who can do a piece of work (give us fun) through the way other actors treat it. This happens when one of the family members says, "Let's play Quackle, its fun. We always laugh so much when we play it." (quote from the daughter from the case).

The game is put forward as a solution and as the obligatory passage point (OPP) to social entertainment. The solution simultaneously suggests roles and organizes relations, i.e., a specific network where the family members will become game players, and the living room table and chairs will facilitate that the family can sit close together. No less important is it that the games will establish equality between the players, regardless of age and family position.

In the next moment, the interessement, which actually takes place in parallel with the problematization, the family members are convinced the proposed solution is the right one, and barriers for alternative solutions to the problem are added. One of the things that are cut off is television, a frequently used source of entertainment in the family, when one of the adults says: "We shouldn't watch television, we always do. We should do something together instead." (quote from the episode).

Enrollment is the third moment where the players are enrolled and this entails that we must accept the roles of participants as players of Quackle! and accept the terms of the game.

In the last moment of translation, mobilization, the solution is executed, when the family members sit down with the game and start playing. If the mobilization works and translation process is thus successful, then it becomes possible to experience fun and laugh together. This is exactly what happened to the family via the interaction with the game, which created a lot of laughing, especially when the parents made weird sounds.

The game re-organizes the family's social connections and in so doing builds a new network of actors and agency. The game is what Latour has named a "mediator" that "transform, translate, distort, and modify" relations [11]. But the game does more than alter the social relations. It mediates the body and mind of the individual players. In the following we will address how Quackle! accomplishes the mobilization of the players physical and cognitive abilities.

V. WHAT THE GAME DOES

A game cannot do much itself, but is dependent on other actors, and this is, of course, especially true for board games. Nevertheless, games have agency that makes game players act in a manner they would not have acted without the game. In that sense, the game "does" something in line with Latour's concise statement on what defines an actor: "anything that does modify a state of affairs by making a difference is an actor [...]" [4].

Latour stresses that, when we are studying a network in ANT, we are focusing on the circulation between the connections that make up the network [16]. When we look into the Quackle game, we are looking at how agency is floating between the involved actors, which we will try to demonstrate through an analysis of a play scenario.

First, the scenario in bullet points of the family playing the game:

1) The game is placed on the table and the players sit down around it.

2) The game is opened, and the game elements are displayed. There are animals, barns, and cards and a cloth bag.

3) The animals are hidden in a cloth bag and all players get a barn.

4) Each player pulls an animal from the cloth bag: Player 1 gets a snake, player 2 a dog, player 3 a donkey and player 4 a frog.

5) After all animals and sounds have been introduced, they are stored out of view in the barns.

6) The cards are shuffled and dealt.

7) Everyone is ready and turn their first card.

8) A horse, a cow, a duck and a pig is turned, so there is no match.

9) Next cards are turned: a snake, a pig, a frog and an owl appears, still no match.

10) The third cards are turned: A mouse, a donkey, a rooster and an owl appear.

11) The game gathers speed and the cards are turned a bit faster.

12) The fourth card is turned: a cat, a dog, a cat and a frog.

13) Player 1 shouts "Qu..iau" [sounds a combination of a frog sound and a cat sound] and player 3 "Vu..sh"[a combination of dog sound and snake sound] followed by a grinning "Oh no, uh" and finally player 1 says "Miau" just before playing 3 said "Sssshh".

14) Player 3 must gather player 1's card and the game continues.

This is the basic structure of the game, which continues in a similar manner for a long time (about 30 minutes) before a player wins.

Points 1 and 2 are of practical character, but they help to create the framework for what is going to happen. Thus, the following activities are framed and the game's inscription starts to become clear, especially in the form of the rules. The agency is still with the players. This is also the case in point 3, but here the game starts to gain agency. It starts to have an effect on the players, as it prescribes their actions in the next steps.

Our observations show that at the same time the players build up anticipation about what is going to happen, which is seen by the body movements and heard by the tone and pitch of voices. This anticipation started when the players accepted the game as an OPP. It was especially noticeable in point 4 and 5 where the joy of hiding the animals in the cloth bag and pulling one provides a form of excitement that is particularly evident in the youngest child. Thus, we see here that the agency is distributed to the game as a kind of pre-disposition of body and mind [5].

In point 5, the players need to remember all the animals, the other players have. The individual player has to establish links between the different animals, the barns and the players around the table. In point 7, the number of links is expanded by the creation of connection to the cards and in point 9, the game is made even more complex as more animals are introduced and it makes it harder to remember the animals hidden in the barns, which is of course part of the game designers' inscription.

We continue to point 13, where we see the first match of cards. When this match appears, a special script appears which is part of the inscription of the game. The script forces the player to act as prescribed by the game rules and thereby it functions as a type of mechanism that governs players' actions. The mechanism *re-organizes* the connection between the players body and their cognition in a special way by means of rules and materials (cards, animal figures, barns) and in this manner the game utilizes the faculties of the player. As mentioned earlier, the player is driven to make mistakes when pronouncing words, and it is this "drive" that shows an agency from the game.

What the game does can be described as follows: First, it mobilizes the individual players memory, but overstates the demands of remembering. There is a wide range of images, sounds, figures and places that are in play, and the player will have to revive all of these objects and connections when the match of card appear. There are different animal figures and their sounds to choose from, and several sounds become actualized, before the players say the correct sounds.

Second, the game cuts across the usual connection between the players mind and body. In point 13 it is clear that the game disrupts the normally well-controlled connections between the player's cognitive ability and their ability to control their tongue. The inscription provides a procedure for a specific requested response to certain signals where the player has to use specific cognitive functions, i.e., perceive, remember, associate the images and sounds as well as mobilize the organs of speech, and it all has to happen as quickly as possible. It is a simple task that the players do not usually have problems with, but by adding a wide range of signals in the form of different images and sounds, and by forcing the players to compete with others, the result is that cognitive and bodily functions responds in an incorrect manner, and the players end up saying the wrong sounds. The game has, in a way, taken over body and mind.

The case of playing Quackle! is an example of a translation process in action, where agency is delegated to a network. The case is also an example of how such a network is comprised of human, material, and social actors. The translation is only happening because the players have allowed themselves to be enrolled as players and fulfill their roles, using the material, and following the rules and thereby delegate agency. In return they are entertained.

B. *Playing a computer game*

Earlier, we stated that we consider computer games to be games before anything else. Thus, our thesis is that computer games do something to the players when played, just as the case of Quackle! What we have attempted until now is to establish a framework for analyzing what games do, and, in the following, we will briefly show how the framework could be applied to computer games.

The setting, which we observed, are three boys 12, 12 and 14 years old playing Grand Theft Auto V (GTA) on a Playstation 3. Grand Theft Auto has become very popular with its mixture of racing and adventure, where the players can follow a story already inscribed in the game, but they can also just go racing around in the game city.

The boys take turns at controlling the game, while the two others comment and talk about what is happening. In one scenario, the 14 year old is controlling the game. He gets an assignment from the game where a tough looking guy on the screen tells him that he needs to win a race with a computer-controlled opponent to progress. Then the game begins.

The setting we are analyzing is a network that consists of the interior (couch, table, etc.), the Playstation (consisting of screen, game console, controller and DVD), the three boys, and the game. The game itself consists of multiple actors of which some are activated coupled with the other actors of the network. We do not have room here to analyze all actors and possible networks the game can initiate and will only take a short look at how the game impacts the players' bodies.

When playing, the boys have to follow the rules of the game. They are complicated, but for our example here we can just point to the traffic rules in the game and how the car is driven via the controller. In the same manner as in a real traffic system, the player has to delegate agency to the system. Just as in the real traffic, there is police in the form of multiple cars and helicopters, roads, houses, pedestrians and the normal traffic on the road, which has to be avoided during the race. All of these actors become active as the boy starts the race, which lasts for a few minutes.

It is apparent how the game influences the player's body. First, of the boy presses hard on the controller and swings it forward, and the next second he and the controller are leaning heavily to the left side, almost leaning into one of the other boys. Next second all of the boys shout “Wow, that was close!”, while they all jump a little in the couch. At the end they are all standing up and leaning forward and to the side as they follow the movements of the car on the road it tries to follow.

If we look at this scenario as a translation, we can see the problematization is set forward, the boys need to win the race and this is also the OPP. In the interessement, the game builds on the fact that the boys are already enrolled in the game (emerged in it) and thus they need to progress to keep playing. The enrollment is made more stable by the use of a character in the game and adding a storyline to the race (why they have to win), thus agency is transferred to the game. This also builds up the tension for the next moment, where the boys are mobilized to play. The term “boys” indicated that all three boys participate even though two of them don’t control the game.

When the race begins, the boy controlling is leaning forward and swinging to the side with his body. This is where the game uses some of its agency, and the bodily action of the player shows that the game is mobilizing the player’s ability. In our observations, we saw this again and again, the players could not help it but move their body to the side as they turned a corner, even though in this game it wasn’t needed, as the controller doesn’t react to it.

The game further uses its agency when it makes the boys shout and jump. This happens as the car almost hits a wall that would have crushed the car, and made them lose the game. This kind of danger is present all the time in the race. Here, the game is exercising its agency by using the player’s body and mind, including his imagination that allows him and the other boys to experience danger, which in the real world would have produced fear, but in the framework of the game produces excitement.

VI. Conclusion and Future Work

In the introduction, we stated that games from our point of view could be regarded as actors because they function as organizers of other actors. Following Latour, quoted above games are actors because they make a difference, not because they are human or non-human, social or material. We have tried to show how such “difference” is created when games do something with players. This view represents an understanding of interaction where the subject-object dichotomy is dissolved and agency is distributed in a process of reorganizing, recreation and modification of actions in networks, which even stretch into the mind and body of the individual player and take advantage of abilities and faculties.

If one accepts this way of viewing, this has implications for game design, because design is not just a question of creating game worlds and interfaces, but a question of how to design social actors that can take agency and thereby initiate and guide the building of social networks, which can bring human and non-human actor to act together in such a way that the players can achieve an experience, they find pleasant, joyful, funny or alike. As we have tried to point out, this does not only involve organizing social relations, actions and material, but also requires utilization of the player’s abilities, for instance of physical and cognitive kind. We believe game design should be done on the basis of knowledge about how human abilities can be organized and influenced, including knowledge of the abilities of different user groups. In the analysis, we showed how the games orchestrate actions by humans and non-humans that resulted in experiences the players find engaging, joyful and entertaining. From our point of view that is prototypical examples of what games do. They organize the acting of actors in order to achieve certain kind of experiences. Through the inscription the designer assigns agency in such a way that the game can take advantage of the characteristics of the human players. The games are examples of how the design render agency to a non-human object, and this object they perform a job by getting the players to do a job.

A. *Future work*

The main theme of this paper has been to establish an understanding of what games do in the perspective of ANT. We believe that ANT is beneficial when we look into computer game design. While it can seem trivial that games do something to users, it is highly important for game designers to understand how games do this. We have demonstrated that using ANT as a tool for analysis can give us a new understanding of the interaction between games and users. We believe that game designers can advance interaction design by “following the actors” and by understanding how agency in games works. We are fully aware that our analysis has shortcomings due to only covering two games and, thus, only a few examples of the kind of actor network, which creates play. There are numerous other examples of this kind of network operating in many different ways in games. Future work should focus on identifying, characterising, and possibly systemizing actor networks in different games.

References

[1] P. Dourish (2004): “Where the Action is – The foundation of Embodied Interaction”. The MIT Press, Cambrige

[2] Valve Corporation (2011). *Counter Strike*. Washington:Valve Corporation

[3] Blizzard Entertainment (2013): *World of Warcraft*. Irvine: Blizzard Entertainment

[4] B. Latour (2005): *Reassembling the social: an introduction to Actor–network theory*- Oxford: University Press, 2005.

[5] A. Yaneva (2009): "Making the Social Hold: Towards an Actor-Network Theory of Design." Design and Culture no. 3, 2009, pp. 273-288.

[6] M. Cypher and I. Richardson (2006): An actor-network approach to games and virtual environments. In CyberGames '06: “Proceedings of the 2006 international conference on Game research and development“, 2006, pp. 254-259.

[7] K. Kallio, F. Mäyrä, and K. Kaipainen (2011) “At Least Nine Ways to Play: Approaching Gamer Mentalities” Games and Culture, vol. 6, no. 4, 2011, pp. 327-353

[8] U. Plesner (2013) “Researching Virtual Worlds: Methodologies for Studying Emergent Practices”. New York,

Routledge, (Routledge Studies in New Media and Cyberculture, Vol. 14).

[9] K. Salen and E. Zimmerman (2004): “Rules of play: game design fundamentals”. Cambridge, The MIT Presse.

[10] M. Callon and B. Latour (1981): Unscrewing the Big Leviathan: how actors macrostructure reality and how sociologists help them to do so. In K. D. Knorr-Cetina and A. V. Cicourel (eds.) “Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-Sociologies”. Boston, Routledge and Kegan Paul.

[11] B. Latour (1996) “The Trouble with Actor-Network Theory”. In F. Olsen: “Om aktor-netværksteori. Nogle fa afklaringer og mere end nogle fa forviklinger” Philosophia, Vol. 25 N. 3 et 4, 1996, pp. 47-64.

[12] B. Latour (1992) “Where are the Missing Masses? The Sociology of a Few Mundane Artifacts”. In W. E. Bijker and J. Law (eds.): “Shaping Technology/Building Society”. Cambridge, The MIT Presse.

[13] B. Latour (1995) “A Door Must be Either Open or Shut: A Little Philosophy of Techniques”. In A. Feenberg and A. Hannay (eds.): “The Politics of Knowledge”. Bloomington, Indiana University Press.

[14] O Hanseth and E. Monteiro (1998) “Understanding Information Infrastructure”, University of Oslo (http://heim.ifi.uio.no/oleha/Publications/bok.html, [retrieved: January, 2014])

[15] M. Callon (1986) “Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay”. In J. Law (eds.), “Power, Action and Belief: A New Sociology of Knowledge”. London, Routledge & Kegan Paul.

[16] B. J. Kraal (2007) “Actor-network inspired design research: Methodology and reflections”. In Proceedings Internationl Association of Societies for Design Research, Hong Kong, 2007, pp. 1-12.

[17] J. Rhodes (2009) “Using Actor-Network Theory to Trace an ICT (Telecenter) Implementation Trajectory”. Information Technologies & International Development, Vol 5, Issue 3, 2009, pp. 1-20.

[18] M. Akrich (1992) “The De-scription of Technical Objects”. In W. Bijker and J. Law (eds.) “Shaping Technology/Building Society: Studies in Sociotechnical Change”. Cambridge, The MIT Presse.

[19] J. P. Spradley (1980) “Participant Observation”. Orlando, Florida, Harcourt College Publishers, pp. 58-62.

[20] P. Atkinson and M. Hammersley (1994). "Ethnography and Participant Observation". Handbook of Qualitative Research, pp. 248–161.

[21] U. Flick (2006). “An Introduction to Qualitative Research”. 3rd ed., London, Thousand Oaks.

[22] Algaspel (2011). Quacklemanual.

 ISBN: 978-1-61208-325-4

Towards Essential Visual Variables in User Interface Design

Johanna Silvennoinen
Department of Computer Science and Information Systems
University of Jyväskylä,
Jyväskylä, Finland
e-mail: johanna.silvennoinen@jyu.fi

***Abstract*—This paper focuses on visual variables in user interface design from the user perspective. Visual design of user interfaces is essential to users interacting with different software. The study is conducted with 3E-templates for users to express their impressions by writing and drawing regarding visual website design. The data is analyzed with qualitative content analysis through interpretation framework. The results of this study provide new insights into user-centered visual user interface design. The results indicate which are the most essential visual variables in user interface design and therefore should be emphasized in user-centered design that promotes positive UX and, thus, benefit user-centered visual UI design. The results of this study are beneficial for researchers in the field of visual aesthetics in human-computer interaction and to user interface designers.**

Keywords-visual user interface design; visual variables; user-centered design; human-computer interaction

I. INTRODUCTION

Visual design of user interfaces (UIs) is significant to users interacting with different software. In addition to the traditional efforts in human-computer interaction (HCI), current and future UIs need to be developed in a novel way to match the potentials offered by the newest computing technologies and the users´ requirements based on their visual experiences. However, the snowballing research of UI design has, until now, largely left the study of visual variables in UI design aside as means to enhance user experience (UX). Current research concentrating on visual UI design in the research area of visual aesthetics in HCI has mainly focused on high-level attributes [1]. The high-level attributes, which have been in focus are, for example, unity and prototypicality [2], novelty [3], typicality and novelty [4]. Visual aesthetics in HCI lacks knowledge of low-level attributes, i.e., visual elements, such as color, size, and balance [5]. In this paper, a qualitative approach is adopted in order to increase knowledge of visual elements in UI design.

In addition, from a user-centered viewpoint, the rapid development of UI technologies demands clarification of how visual variables should be utilized in UI design to promote positive UX. In this study, visual representations that draw on the theories of two-dimensional pictorial elements are utilized as theoretical vehicles to create novel insight into visual UI design. The dispensation of visual theory is in that it provides knowledge of exact visual variables, such as size, value, hue, orientation, contrast, texture [6], shape, proportion, and position [7] and, for instance, form, and expression [8], that are seen as visual elements inherent in contemporary UIs. Mullet and Sano [7] refer to visual language in designing visual UIs. Through visual language, based on visual elements, UIs communicate to the users. Visual language of UIs is divided into visual UI design factors, which are visual characteristics (shape, color, position, texture, size, orientation, etc.) in a specific set of design elements (point, line, volume, plane, etc.) and the factors by which they relate to each other; such as balance, structure, proportion, and rhythm.

These elements facilitate in finding the essential visual elements that should be applied intentionally to modern UI design in a manner that enhances interaction between UIs and users. Further, visual theory serves as a starting point in the study because aesthetic impressions conveyed by visual representations are the best way of evoking UX [9] that may promote positive feelings on UIs [10] due to the emotional nature of an aesthetic experience [11], [12].

In addition, the improvement of visual aesthetics of UI design will also enhance the usability of the product, by promoting visual organization, clarity, and conciseness of UIs [7]. More profound understanding of visual variables in UI design from user-centered perspective contributes to enhanced visual usability of UIs [13]. Moreover, in the current era of visual UIs, usable designs need to highlight aesthetic expression as meaningful presence for users instead of just providing designs from the functional tool perspective [14]. Clearly, research on users' experiences on specific visual variables in UIs is needed.

Aesthetic qualities are currently emphasized in designing software for HCI. According to Tractinsky [1] usability experts and designers came to the conclusion that these two aspects of design, visual aesthetics and usability, could and should coexist in the same context of use. Before this shift in the first decade of 21th century, visual aesthetics and usability were often seen to have a contradictory relation, in that, when one was emphasized the other one was automatically omitted. The shift has emerged mostly because recent research corroborates a positive correlation between aesthetic and usability principles [9], [15], [16], [17].

Yet, often in the software development process, only the implementation of software (e.g., programming) and the traditional, general usability design [18] are considered as the most important parts of the process [19]. In reality, users'

 ISBN: 978-1-61208-325-4

needs are not taken into account in the software development process [20]. The essential deficit in UI design lies in the absence of visual UI design specialists in the development process.

This paper focuses on studying the visual elements as variables in UIs in order to provide usable insight into the evaluation and design of visual variables in UIs that promote UX and thus, to benefit user-centered visual UI design. This study focuses on shedding light into the following research questions: *which visual variables contribute to positioning of the viewer as UIs´ ability to communicate the whole content in a visual manner as quick and easy to grasp? What are the essential visual variables in modern UIs´ from a users' point of view?*

The rest of the paper is organized as follows. First, the method for the study is described. Second, the analysis of the data is presented. Third, the results of the study are discussed and finally, conclusions with future research and limitations of the study are presented. The results of this study will provide novel insights into what are considered as the most essential visual variables in UI design. Also, the results provide knowledge of what is considered important in visual UI design for future programmers, engineers, and practitioners in the field of information technology.

II. METHOD

When people encounter visual representations, such as visual UIs, numerous visual elements are encountered [21]. People may experience visual elements as both explicit and simple, but also as elements involving interpretation. Therefore, the research method should facilitate the combination of the analysis of qualitative and quantified issues. As Collier [22] points out, a studied phenomenon must first be examined without a precocious descent into analyses, maintaining a focus on pre-existent structures and points of interest. In this study, the data is collected and analyzed with a mixed methods approach [23].

First, a data driven qualitative analysis is conducted to find and describe the UI elements depicted by the informants after having looked at the selected UIs. Second, the elements found and defined were quantified in order to find the visual variables that are considered important in visual user interface design, particularly in layout design. The analysis was conducted with these two procedures, which supported the analysis from two different viewpoints.

The quantification of the explicitly written words representing specific objects is important, but it also relates to the qualitative procedures. Krippendorf [24] emphasizes the meaning of context in content analysis. Texts and images are always produced in some specific cultural context and they also refer to wider cultural context. Regarding content analysis, this aspect is considered by first deploying the interpretative viewpoint as an independent phase before quantifying the found elements.

The methodological decisions for the study are influenced by the nature of the visual viewpoint: instead of directly analyzing visual UIs, the data is comprised of participants' descriptions of those visual UIs. Therefore many methods, for example visual anthropology [22], semiotics and iconography [25] as well as social semiotic visual analysis [26], are not applicable because they assume that the data along with the object of analysis are visual images. Furthermore, even though the data collected by other methods (such as eye tracking) could enable the extraction of specific points of attention, it would be impossible to analyze which particular element draws attention; color, form or, for instance, a tension between elements.

The data collection was conducted in the University of Jyväskylä in the "The User-centered software development: Designing, realizing and testing the visual appearance and usability of user interface" course, held in the Department of Mathematical Information Technology. The course included 10 lectures and 5 demo exercises. Three lectures dealt with visual UI design. The first lecture of visual design concentrated on user-centered visual UI design, the second dealt with layout design and composition and the third with color design. The data was collected with the 3E (expressing emotions and experiences) -template [27], [28] from the lecture dealing with layout design. The 3E-template was selected as a data collection method to allow respondents to express their thoughts both verbally and non-verbally: by writing and drawing (Fig. 1).

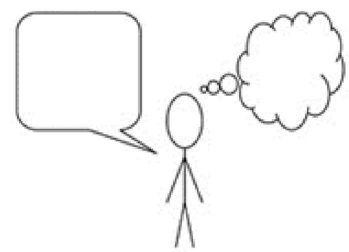


Figure 1. The 3E-tempalate.

From the beginning of the lecture, participants were familiarized with the design process in general, the evaluation methods of visual UIs, pictorial elements, such as lines with different characters, visual rhythm, dynamics, balance, tension, contrast and Gestalt laws. After this part of the lecture, the participants were introduced to the 3E-template and they were asked to write and draw their thoughts and impressions of two still pictures of web pages, with a unique focus on the compositional elements.

The participants took approximately 20 minutes in answering the two templates. The research data is comprised of the written and drawn reflections and interpretations on the compositional aspects of example web pages. A total of 100 templates were answered; 50 templates times two visual web pages.

The common factors between the participants were that they all were in the same auditorium, heard and saw the same content of the lecture, and were all introduced to the web pages as well as to the template at the same time with the same advice, and were allotted the same amount of time for answering the template. In addition, all the participants were university students, mainly from the faculty of information technology. The group of participants included both novice and expert users of information technology. Of the participants, 18.5 % were female and 81.5 % male. The age of the participants ranged from 19 years to 50 years. The average age of the participants was 25 years, standard deviation of the age was 5.8. The circumstances of the data collection were designed similarly for all the respondents.

The objects for the compositional reflection were two web pages (Fig. 2) from the CSS Zen Garden web page gallery, a web page created with CSS-based design [29]. All the web pages have the exact same content but altering visual appearances. The CSS Zen Garden web pages were scoured in order to find two example layouts that would differ from each other. Especially with regards to the amount of elements that divide the surface; diagonal and horizontal lines, and the overall use of space.

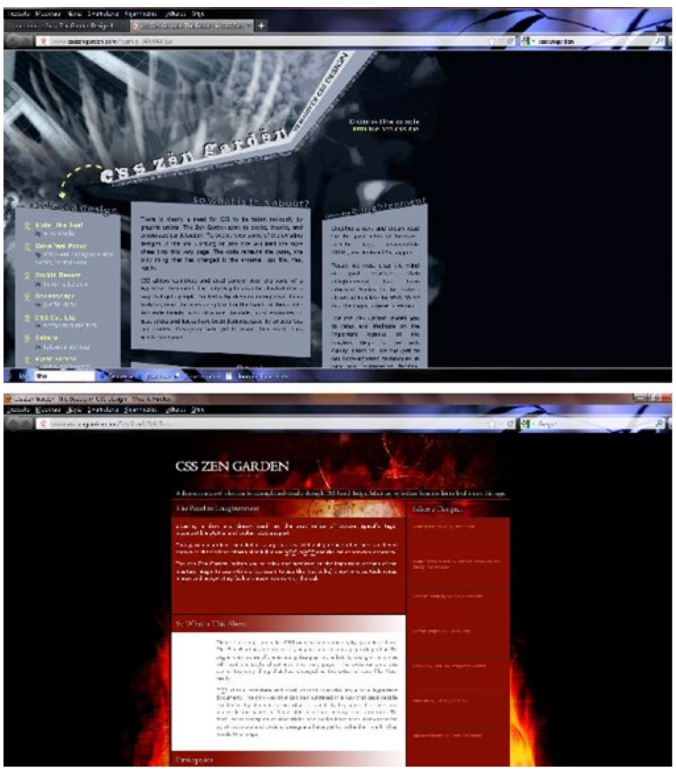


Figure 2. The two UIs used in data collection.

All the CSS Zen Garden web pages were looked through and opened in new tabs. The web pages were first divided into two categories and then compared step by step, finally resulting in two example web pages.

Emphasis on a choice of web pages according to their differentiating elements was made in order to provoke participants towards a comparative analysis between web page layouts. Bell [30] also emphasizes that evaluations with content analysis are often comparative. The web pages were therefore selected with a comparative setting, regarding the pages' differences in constructing visual elements. This was achieved in the study by using two different web pages with similar content but differing visual appearances.

III. ANALYSIS

The data analysis proceeded as follows. First, the data was observed as a whole by reading the templates. The purpose was to first focus on the visual variables in a neutral context in order to gain an understanding of the visual elements that are seen as important in visual UI design despite of the emotions they might evoke. Therefore, emotional aspects of the visual variables are not studied in this paper. The human figure in the templates was only used in a few papers to illustrate emotions by drawing facial expressions. Almost none of the drawings emphasized the written content or the facial emotions of the figure, but brought to attention something without a clear connection to the written content. Due to these characteristics of data, the focus of the analysis was on the written texts describing the visual elements.

The aim was on finding the visual variables that have drawn the most attention and can be seen to have importance due their frequent emergence and contentual significance. After becoming acquainted with the data, a data-driven interpretation rule was developed, which was used to assist in the analysis. The interpretation rule included the items, which directed the focus on a conceptual level during the data analysis, in interpreting and comparing interesting insights within the data. The interpretation rule consisted of compositional interpretation [31] and visual variables in UI design [7].

Compositional interpretation refers to describing the appearance of images with a detailed terminology. This form of visual analysis requires contextual knowledge of art work and a particular way of looking at images ("the good eye"), which is not methodologically explicit, but functions as visual connoisseurship and is a specific way of describing images. Compositional interpretation focuses on the image itself by trying to comprehend its significance mostly by focusing on its compositionality. Interpretation does not focus on "external factors" such as, what kind of messages the image sends and does it have some functional meaning.

The terminology of compositional interpretation includes several components. First, the content: what the image actually shows. The second component is color, which is more specifically defined with concepts of hue, saturation value and the harmony of color combinations. Thirdly, spatial organization, which includes volume, lines, static and dynamic rhythm, geometrical perspective, logic of figuration (how the elements of a picture offer particular viewing position outside the photo) and focalisers (the visual organization of looks and gazes inside the picture and in relation to the viewer's gaze). The fourth is light, what type of light is represented and from what sources of light. The last component is expressive content, which describes the "feel" of the image combining the effect of the subject matter and visual form.

Compositional interpretation approach is established in Art History and is usually used in studying paintings [31]. Visual UIs can be comprehended as paintings or, in a more general viewpoint, as any two-dimensional representation that is constructed with pictorial elements, such as lines and volumes. Compositional interpretation can, therefore, also be extended to analyzing visual UIs.

Mullet & Sano [7] present visual UI design factors, which are shape, color, position, texture, size, orientation, point, line, volume, balance, symmetry, scale, contrast, structure, proportion, rhythm and position. The emphasis on visual factors in UI design is to provide insight into designing good visual usability and effective visual communication. They also point out that negative space and grouping are important components in visual UI design.

The interpretation rule combined the discussed visual elements. The elements were derived from the two approaches described above and created the content of the interpretation rule. The interpretation rule included, but was not restricted to, the following components: color (hue, saturation, value, harmony of color combinations), spatial organization, geometrical perspective, volume, lines, points, size, texture, shape, static and dynamic rhythm, orientation, balance, symmetry, scale, structure, proportion, negative space, grouping, position, figuration logic, focalisers, contrast and light.

From a compositional interpretation, the expressive content component was excluded from the interpretation rule because of its emphasis on subject matter evoking emotionality which is not in the focus in this study. The templates that comprise the data were mainly used by the participants to express their impressions of the UIs in written form. Examples of written expressions in the speech and think bubble are presented in Fig. 3 below.

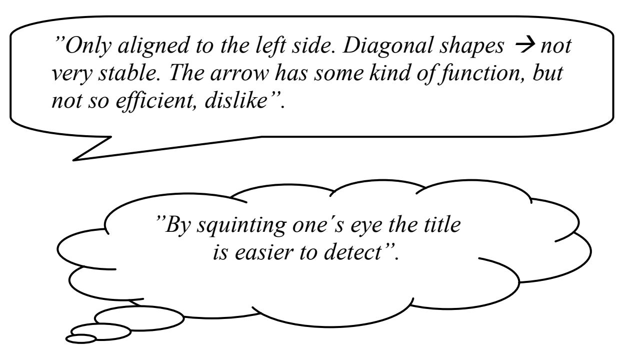


Figure 3. Examples of written expressions.

The interpretation rule functioned as a ‘theoretical lens’ in the analysis. The lens was constructed based on the elements that were included in the rule. The analysis required accuracy and concentration in detecting tiny nuances in finding the relations between the different visual elements, which required close attention in detecting how many altering ways there are, for instance, to describe the use of space in UIs. All the 100 templates were analyzed by writing down all the described elements in the templates and by counting the frequency of their occurrence.

The second phase was to create categories of relations and then to critically combine several different categories of relations between elements, into main categories. For instance, from the example described above the first observation about left side alignment is related to the spatial organization as is the second notion of diagonal lines. The stability refers to symmetry and the role of the arrow is related to functionality. The interpretation rule guided the detection of written reflections of the described elements. For instance, the interpretation rule did not include usability, which emerged from the data as a connective factor between relations of different elements, especially in terms of visual usability [13] of the UI. In many templates think bubbles were used to express additional reflections, mostly about the supposed functionality of the UI.

In the data, spatial organization was reflected in detail and visual elements referring to this category were most often mentioned. Observations of balance and imbalance were seen in relation to symmetry. Justification and alignment to the left or right were often in the same context as the sensation of space. The use of negative space was taken into account, but overall was not in focus. Another important relation was often found between the continuity, groupings and togetherness of the whole.

Even though the participants were instructed to only use the template for reflecting the impressions of the compositional elements of the layout, a lot of attention was paid towards contrasting colors and their role in visual usability. In relation to contrast most of the remarks were about contrasting colors, especially between the texts and background, which was, moreover, attached to readability. Size was only discussed in the context of the font size.

The position of the viewer (including figuration logic and focalisers) was seen as a visual strategy that guides the viewer’s gaze in the UI. The shape of an arrow along with horizontal, diagonal and vertical lines, were especially emphasized and seen as visual elements that guide the gaze forward. There was a strong emphasis on describing how the visual elements function and how they support visual usability and interaction, especially the ways by which visual elements direct attention in the UI towards the important areas. Without knowledge of the context and aid of the interpretation rule, the data driven qualitative interpretation would have been problematic.

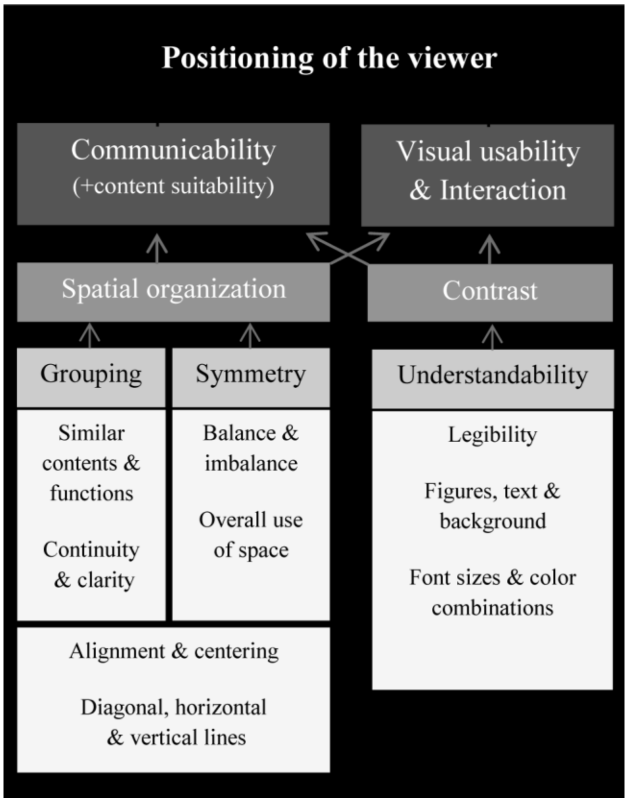
IV. RESULTS

The analysis is reflexively highlighted and evaluated in the following sections. Implications for future research and limitations of the study are also discussed. Based on the results, insights into the essential visual variables in UIs are represented in Table 1: Essential visual variables contributing to positioning of the viewer.

Positioning of the viewer refers UIs ability to communicate the whole content in a visual manner as quick and easy to grasp. Positioning of the viewer gathers all the results under one definition, and functions as the main viewpoint in interacting with visual UIs from user-centered perspective. Positioning of the viewer was emphasized regarding the communicative ability of the visual variables [7], and in relation to the UIs content suitability. The overall visual impression of the UI´s appearance should be suitable for the context for which it is designed.

 ISBN: 978-1-61208-325-4

TABLE I ESSENTIAL VISUAL VARIABLES CONTRIBUTING TO POSITIONING OF THE VIEWER



Besides communicability, positioning of the viewer was approached from the viewpoints of visual usability and interaction. Visual usability and interaction include the ways how different visual elements guide the interaction and usability. Users appreciate UIs that fluently communicate the content in an understandable visual form. Thus, in designing visual UIs the power of the visual elements to guide users´ attention need to be taken into account.

Positioning the viewer was discussed in relation to visual variables contributing to spatial organization and contrast. Spatial organization and contrast were seen to apply both to communicability and to visual usability and interaction. Spatial organization was emphasized by focusing on grouping and symmetry. Grouping and symmetry were seen as the two primary visual variables essential to spatial organization of visual UI elements, and in creating an impression of the UI as a consistent totality, in which visual variables contribute to the impression of the UI as a whole.

Grouping of similar contents and functions were seen as substantive factors creating continuity and clarity. Spatial organization through balance and symmetry was seen important regarding the overall use of space. Grouping and symmetry were further discussed in terms of aligning or centering the content, and were seen constitutive factors influencing both grouping and symmetry. Diagonal, horizontal and vertical lines were highly emphasized as visual variables that strongly affect the positioning of the viewer and lead the interaction as the main low-level variables.

In addition to spatial organization, contrast was seen as influential variable. Functional and communicative impressions of the visual elements were in focus regarding understandability and legibility of the content. Contrast between figures, text and background was conveyed through different font sizes and color combinations. These ways in creating contrast between visual elements in UIs were often mentioned, especially concentrating on the contrast between background color and the size and color of fonts. Moreover, designing contrast between these elements was highlighted due to its ability to promote effortless visual usability, which leads to better UX in interacting with visual UIs. Essential visual variables contributing to fluent and pleasurable interaction have a strong impact on users´ perceptions and interpretations of the UIs´ communicative abilities and visual usability. Therefore, in designing visual UIs, the essential visual variables should be noticed as factors contributing to position the viewer.

V. CONCLUSION AND FUTURE WORK

The results of this study provide novel insights into what are considered the most essential visual variables in UI design that contribute to communicability, visual usability and interaction aiming to position the viewer. Spatial organization and contrast, with the lower-level variables which they are based, such as diagonal lines and font colors, are essential to users in visual UI design that promote fluent interaction and visual usability. Characteristics of lines have an important role in emphasizing visual usability and communicability of UIs. Spatial organization through balance and symmetry contribute to the overall impression of the UI, and grouping similar objects contributes to understandable and effective design. Contrast between figures, texts and backgrounds enhance visual usability and interaction with the UI. In addition, through contrast UIs can be designed to communicate the content in an effortless manner, by grasping the content with one glance. Communicability of the UI also refers to the content´s suitability in relation to the context.

Without contrast between visual elements, UIs could not communicate the content to the users interacting with them. This notion emphasizes the essence of visual variables contributing to fluent HCI. UI´s ability to position the viewer through visual variables is essential for pleasurable interaction and efficient visual usability leading to good UX. Moreover, the results of this study provide knowledge of what is considered important in visual UI design among future programmers, engineers, and practitioners in the field of information technology.

Future research focuses on proceeding from specific contexts towards the discussion of a more general understanding of the studied phenomena. Data collection could also be conducted with participants, with differing backgrounds in order to get more insight in dealing with the importance of different visual variables. In addition, participants could vary more according to their cultural backgrounds in order to provide understanding of the essential visual variables in different cultures. Moreover, the UIs of this study could also be analyzed through different data collection methods, such as questionnaires, in order to test the validity of 3E-template and to compare different approaches in relation to the subject of study.

Visual design of UIs is not universally understood and appreciated similarly in different cultures. Different meanings attached to, for instance, dimensions of visual space are highly influenced by the writing and reading direction [32]. Therefore, due to the context of this study, the results can be applied to Western culture. In addition, this study was conducted in relation to visual web site design and therefore, might not be applicable to other visual UI design contexts, such as mobile UI design.

ACKNOWLEDGMENT

I thank COMAS (Doctoral School of the Faculty of Information Technology in University of Jyväskylä) for making this research possible.

REFERENCES

[1] N. Tractinsky, "Visual aesthetics: in human-computer interaction and interaction design," In Encyclopedia of human-computer interaction, M. Soegaard, and D. Rikke Friis, Eds. Aarhus, Denmark: The interaction-design.org foundation, 2012. Available online at http://www.interaction-design.org/encyclopedia/visual_aesthetics.html.[retrieved: 01/2014]

[2] R. W. Veryzer and J. W. Hutchinson. "The Influence of unity and prototypicality on aesthetic responses to new product designs," Journal of Consumer Research, vol. 24(4), pp. 374–385, 1998.

[3] W. –K. Hung and L. L. Chen, "Effects of novelty and its dimensions on aesthetic preference in product design," Int. J. of Design, vol. 6(2), pp. 81–90, 2012. Available online at http://www.ijdesign.org/ojs/index.php/IJDesign/article/viewFile/1146/474 [retrieved: 01/2014]

[4] P. Hekkert, D. Snelders and P. C. W. Wieringen, "Most advanced, yet acceptable: Typicality and novelty as joint predictors of aesthetic preference in industrial design," British Journal of Psychology, vol. 94(1), 111–124, 2010.

[5] D. Cyr, M. Head and H. Larios, "Colour appeal in website design within and across cultures: A multi-method evaluation," International Journal of Human-Computer Studies, vol. 68(1–2), pp. 1–21, 2010. Available online at http://www.sciencedirect.com/science/article/pii/S1071581909001116 [retrieved: 01/2014]

[6] M. Wiberg and E. Robles, "Computational compositions: Aesthetics, materials, and interaction design," Int. Journal of Design, vol. 4(2), pp. 65–76, 2010. Available online at http://www.ijdesign.org/ojs/index.php/IJDesign/article/viewFile/757/290 [retrieved: 01/2014]

[7] K. Mullet and D. Sano, Designing visual interfaces: Communication oriented techniques. New Jersey, USA: Prentice Hall, 1995.

[8] L. Hallnäs, "On the foundations of interaction design aesthetics: Revisiting the notions of form and expression," Int. J. of Design, vol. 5(1), pp. 73–84, 2011. Available online at http://www.ijdesign.org/ojs/index.php/IJDesign/article/viewFile/689/319 [retrieved: 01/2014]

[9] N. Tractinsky, A. S. Katz and D. Ikar, "What is beautiful is usable," Interacting with Computers, vol. 13, pp. 127–145, 2000.

[10] D. Norman, Emotional design: Why we love (or hate) everyday things), New York, USA: Basic Books, 2004.

[11] J. Dewey, Art as Experience, New York, USA: Capricorn Books, 1958.

[12] P. Hekkert, "Design aesthetics: principles of pleasure in design," Psychology Science, vol. 48(2), pp. 157–172, 2006.

[13] T. Schlatter and D. Levinson, Visual usability, principles and practices for designing digital applications, MA, USA: Morgan Kaufman, 2013.

[14] L. Hallnäs and J. Redström, "From use to presence: On the expressions and aesthetics of everyday computational things," ACM Transactions on Computer-Human Interaction, vol. 9(2), pp. 106–124, 2002.

[15] N. Tractinsky, A. Cokhavi, M. Kirschenbaum and T. Sharifi, "Evaluating the consistency of immediate aesthetic perceptions of web pages," International Journal of Human-Computer Studies, vol. 64(11), pp. 1071–1083, 2006.

[16] M. Thüring and S. Mahlke, "Usability, aesthetics, and emotions in human-technology interaction," International Journal of Psychology, vol. 42, pp. 253–264, 2007.

[17] A. Sonderegger, G. Zbinden, A. Uebelbacher and J. Sauer, "The Influence of product aesthetics and usability over the course of time: A longitudinal field experiment," Ergonomics, vol. 55(7), pp.713–730, 2012.

[18] J. Nielsen, Usability engineering. San Fransisco, CA: Morgan Kaufman, 1993.

[19] J. Iivari, H. Isomäki and S. Pekkola, "The user – the great unknown of systems development: reasons, forms, challenges, experiences and intellectual contributions of user involvement," Information Systems Journal, vol. 20, pp. 109–117, 2010.

[20] B. Bygstad, G. Ghinea and E. Brevik, "Software development methods and usability: Perspectives from a survey in the software industry in Norway," Interacting with Computers, vol. 20, pp. 375–385, 2008.

[21] H. Zettl, Sight, sound, motion: Applied media aesthetics, Boston: Wadsworth Publishing Company, 1999.

[22] M. Collier, "Approaches to analysis in visual anthropology," In handbook of visual analysis, 4th ed., T. Van Leeuwen and C. Jewitt, Eds. London: Sage, 2004.

[23] J. W. Creswell and V. L. Plano Clark, Designing and conducting mixed methods research, Thousand Oaks, CA: Sage, 2007.

[24] K. Krippendorf, Content Analysis: An introduction to its methodologies. London: Sage, 1980.

[25] T. Van Leeuwen, "Semiotics and iconography," In handbook of visual analysis, 4th ed., T. Van Leeuwen and C. Jewitt, Eds. London: Sage, 2004.

[26] C. Jewitt and R. Oyama, "Visual meaning: a social semiotic approach," In handbook of visual analysis, 4th ed., T. Van Leeuwen and C. Jewitt, Eds. London: Sage, 2004.

[27] M. Tähti and L. Arhippainen, "A proposal of collecting emotions and experiences," vol. 2 in Interactive Experiences in HCI, pp. 195–198, 2004.

[28] M. Tähti and M. Niemelä, "3E - expressing emotions and experiences," Proc. WP9 Workshop on Innovative Approaches for Evaluating Affective Systems. HUMAINE (Human-Machine Interaction Network on Emotion), 2006.

[29] CSS Zen Garden gallery: http://www.csszengarden.com

[30] P. Bell, "Content analysis of visual images," In handbook of visual analysis, 4th ed., T. Van Leeuwen and C. Jewitt, Eds. London: Sage, 2004.

[31] G. Rose, Visual methodologies. An introduction to the interpretation of visual materials. London: Sage, 2005.

[32] T. van Leeuwen and G. Kress, Reading images. The grammar of visual design (2nd ed.), Taylor & Francis e-Library, 2006.

Traffic Light Assistant - What the Users Want

Michael Krause, Alexander Rissel, Klaus Bengler
Institute of Ergonomics
Technische Universität München
Garching, Germany
emails:{krause@tum.de, rissel@lfe.mw.tum.de, bengler@tum.de}

***Abstract*—In a driving simulator experiment, a prototypical traffic light phase assistant is assessed. The main research issue: How would a user customize the system? As a sideline, data is gathered with a special Detection Response Task (DRT), the *Tactile Detection Task (TDT)*, in conjunction with an auditory cognitive task as reference. Recorded gaze data, driving behavior, subjective ratings with a *System Usability Scale (SUS)* and an *AttrakDiff2*-questionnaire are also reported. The subjects were able to customize ten parameters of the traffic light assistant system. The so personalized system configuration showed no great enhancement in the subjective ratings; thus, the later application implementation will include only little configuration features for the user. However, the test persons exhibited a willingness to be informed about speeding by a speed alerting function within the traffic light assistant system. The performance (reaction time) of the TDT is interpreted as a measure for the cognitive load while using the interface. The auditory cognitive task prolonged the reaction times for a tactile detection task more than the traffic light information system. The glance times are in line with current guidlines and the driving behavior shows a potential benefit for safety. Thus, the reported experiment evaluates an interface for use while driving with objective metrics regarding distraction and subjective results related to usability and joy-of-use.**

Keywords-in-vehicle information system*; *IVIS*; *nomadic device*; *tactile detection task*; *TDT*; *glance duration

I. Introduction

In the project KOLIBRI (Kooperative Lichtsignaloptimierung – Bayerisches Pilotprojekt; engl: cooperative optimization of traffic signal control), the Institute of Ergonomics at the Technische Universität München was responsible for the human factor of a traffic light assistant. One of the goals of the project was to provide the driver with information about the state of an upcoming traffic light.

By introducing a traffic light assistant for smart phones, previously installed 2nd and 3rd generation telecommunication networks (e.g., Global System for Mobile Communications, GSM) could be used to transmit the information. Nowadays smart phones are widely used, so there are no extra costs to get an additional display into the car. With a mobile solution, instead of an in-vehicle implementation, there is also no limitation regarding the make, year or brand of the car.

In this project, the traffic lights were equipped with mobile network transponders to send their current state to a central server. The server estimates how the traffic lights will probably act the next time. These estimated switching times can be polled by car systems or smart phones. The devices have to calculate recommendations, based on human factors, and show appropriate information to the driver. Because they are used while driving, special care must be taken for suitability while driving [1].

The system was implemented in two real test fields. The first is in the north of Munich. Over a length of about seven kilometers, seven traffic-light-controlled intersections on federal road B13 were involved. The second test track was a rural road near Regensburg. Over a length of about five kilometers, eight traffic light controlled intersections shared the system.

In order to make a judgment about safety issues, the test track in the north of Munich was modeled for the static driving simulator at the Institute of Ergonomics and was used in an initial subject test [2], [4]. The test persons evaluated five rapid prototyped Human Machine Interfaces (HMI) with standardized subjective questionnaires and wore an eye tracking system. The HMIs were shown on a smart phone and were coupled with the driving simulator. The results led to a favorable HMI within the project, and the gaze behavior showed no critical metrics for this HMI.

In the second driving simulator experiment, reported in this paper, subjects were told to customize the assistant to their needs by adjusting some parameters. The main idea: What aspects must be configurable (e.g., what are potential items on a configuration menu in an application). The test persons also wore an eye tracking system here. In addition, a special Detection Response Task (DRT), the Tactile Detection Task (TDT), was operated. DRTs are currently being standardized and are promising candidates to get objective data for the mostly invisible cognitive load.

The next section presents related work for this paper in the fields of traffic lights assistance and detection response tasks. The method section holds information about the conduction of the experiment. The design of the experiment, technical data of the driving simulator and the traffic light assistant as well as the details of the cognitive task and tactile detection task are reported. The method section closes with the task instruction and demographic data of the participants. The result section shows how the participants would customize the traffic lights assistant and their performance in the tactile detection task. Subjective data from questionnaires compares the liking of the individually customized (HMI) with a default HMI. The section closes with the presentation of glance durations and driving behavior metrics.

II. Related Work

A. Traffic Light Assistance

Early research in the field of in-car traffic light assistance took place in the ’80s in the project *Wolfsburger Welle* from Volkswagen, Germany [5], [6]. At about the same time, Australian traffic engineers experimented with a roadside traffic light assistant along a street in Melbourne [7], [8]. These projects evaluated and identified the benefits for the informed driver, such as fuel reduction. To provide information about the traffic light to the driver, with countdowns, is common in some countries, mainly in Asia. Many people have reflected about this topic, so different solutions for traffic light information can be found in patent classes such as *G08G 1/096*. The advantage of roadside solutions is that everyone can use them. The disadvantage, other than maintenance costs, is an only temporally visibility, or else several must be placed along the road. Another problem is reported by [9]: Counting down the remaining green time at an intersection results in a higher crash risk.

Pauwelussen et al. [10] compared a road side system with an in-car system and found objective reasons for using an in-car system and subjective reason for using the roadside signs. Thoma [11] evaluated various HMIs for an in-car on-board system and proposed a combination that would work with the speedometer. Another project dealing with on-board traffic light guidance is TRAVOLUTION from Audi AG, Germany [12]. A traffic light phase assistant was also included in the German car-to-infrastructure project *simTD* [13]. The German project *AKTIV* built a traffic light assistant on a personal digital assistant (PDA) via WiFi [14], [15]. Another project that used a mobile device for a traffic light assistant is *SignalGuru* [16]. This project heavily relies on the camera of a smart phone on image processing. One radical idea that also involves traffic light assistance is to replace the physical traffic lights with in-car information i.e., Virtual Traffic Light (VTL) [17].

B. Detection Response Tasks (DRTs)

In detection response tasks, the test persons have to react to a continuously repeated stimulus. Typically, this detection task is the ‘probe’ i.e., a measurement tool to asses the demands of another task or combination of tasks, like interacting with a system while driving a car. The prolongation of reaction times and a drop in the rate of successfully fulfilling these detection tasks are potential indicators for the cognitive load. The DRTs are currently being standardized by the *ISO TC22 SC13 WG8*. In former work and projects, the detection tasks had shown the potential to detect cognitive load effects. The use of a vibration stimulus (tactile detection task, TDT) overcomes various disadvantages such as the visibility of a light stimulus under changing lighting conditions and has, if at all, only a weak competition and distraction effect on visual resources. A commendable review of the research on TDT can be found in [18].

III. Methods

A. Experimental Design & Procedure

Each subject drove for each part of the experiment (within design). The test persons first completed a letter of consent and a demographic questionnaire. Afterward they get general explanations about the experiment and the driving simulator. Once they were seated in the mockup car, the gaze tracking systems was calibrated for each person. The subjects drove an acclimatization round without the traffic light assistant and one round with the system (in configuration *complex*). Before and after the core of the experiment, the TDT was carried out alone (single task) for one minute (*TDT_base1* and *TDT_base2*). In the core of the experiment, four parts were completed in randomized order:

- *Baseline*: Driving the simulator without the traffic light assistant and with TDT
- *COTA*: Driving the simulator without the traffic light assistant, but with a cognitive task (COTA) and with TDT
- *HMI complex*: Driving the simulator with the traffic light assistant in a general, predefined configuration and with TDT
- *HMI individual*: Driving the simulator with the traffic light assistant in a personalized configuration and with TDT

Before the *HMI individual* part, the test subject was able to customize the HMI using ten parameters, and was allowed to drive and test the interface as long as needed. After the *HMI complex* and *HMI individual* sections, the subjects filled in a system usability scale (SUS [21]) and AttrakDiff2 questionnaire [22]. The test track (about 7km) was randomly driven in a north-south or south-north direction. A session typically lasted about 90 minutes.

B. Driving Simulator

The simulator track is a model of a real road section of the federal road B13 in the north of Munich (see [2], [4]) which is also the test bed for real field trials in later experiments. The experiment used the institute’s static driving simulator. Three projectors (1400x1050 resolution) show an almost 180-degree front view on 3.4 m x 2.6 m screens. Three other projectors displayed images on screens behind the BMW E64 mockup for the car mirrors. The driving simulation SILAB V3.0 from WIVW GmbH, Würzburg was used together with CarSim V7.11 from Mechanical Simulation, Ann Arbor, as well as an active steering wheel with software from Simotion, Munich. For the eye tracking, the head-mounted system Dikablis from Ergoneers, Manching was used. Gaze analysis was carried out with D-Lab (Ergoneers, Manching) and Matlab.

C. Traffic Light Assistant - Human Machine Interface, HMI

In an initial simulator experiment [2], [4] an appropriate HMI was found to communicate a speed recommendation to the driver via the smart phone (called the *velocity carpet*; see Figure 1, left). If the driver has to move to slowly (below 70% of the speed limit) or too fast to get the next traffic light on green, Figure 1 (middle) is displayed, called *red arrive*. Both conditions (too fast, too slow) are intentionally coded into the same screen, to prevent the misuse of the system. In addition, some new parts were introduced: A circle (left upper corner of Figure 1, left) could provide information about the current cycle state of the next traffic light (called *Heuer traffic light*). A distance bar (right border of the screens in Figure 1) indicates the meters remaining to the next traffic light. A speed alert (Figure 1, right) provides a warning in the event of speeding.

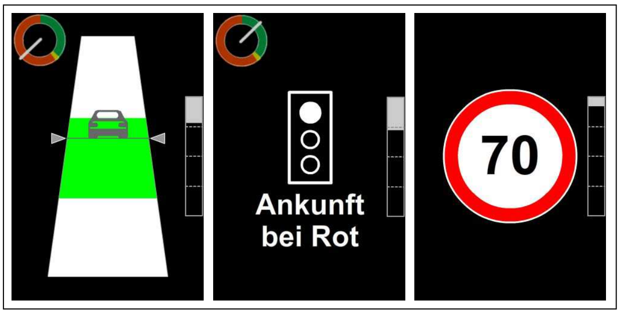


Figure 1. Traffic Light Assistant, Human Machine Interface screens. left:carpet, middle: 'Ankunft bei Rot' = arrival on red, right:speeding

In this experiment, the test persons were able to customize different parameters of the system and test their customized systems directly in the simulator (known as *HMI individual*):

1) the velocity carpet could be switched off (black screen) *[carpet **on**/off]*
2) the velocity carpet could only appear if the distance to the next traffic light is less than X meters, or it can be on all the time *[carpet distance X meters ./. **100%**]*
3) either arrival on red could be displayed, or a blank screen*[arrival on red **on**/off]*
4) *[speed alert **on**/off]*
5) the speed alert could inform the driver if the car is going X km/h above the speed limit *[speed tolerance X km/h]*
6) the Heuer traffic light could be displayed *[Heuer **on**/off]*
7) the distance bar could be displayed *[distance bar **on**/off]*
8) while waiting at a red light, the residual red light time could be displayed *[residual red **on**/off]*
9) when driving through an intersection, the system output could be suppressed (Xing symbol) *[Xroad suppression on/**off**]*
10) when intersection suppression is on it should suppress system output X meters in front and after an intersection [*intersection radius X meters*]

The order of the single parameters for the customization was randomized. The system was also driven in a uniform configuration by all subjects (known as the *HMI complex*). This configuration consisted of the bold options in the previous enumeration and a speed alert tolerance of 0 km/h.

D. Cognitive Task, COTA

For the cognitive task, the program *Cognitive Task 1.0* from Daimler AG (Stefan Mattes) was used with default settings. The program reads a sequence of three numbers (one to nine) out loud, and after a short break a fourth number is announced. The task of the test person is to consistently state whether the fourth number was included among the first. Two examples: Program: *'2,6,9...7'*. Test person: *'No'*. Program: *'7,1,3...1'*. Test person: *'Yes'*. So it is an auditive Sternberg Task. The chosen setting plays a beginning chime sound (about 1s), reads out three numbers (one per second), waits 2s, plays a chime sound (about 1s), announces the forth number (1s), the test person has 3s to answer and at the end the program plays a short closing honk signal and wait 2 seconds before beginning the next sequence. The repeat count was set up to repeat the described 13-second sequence over and over again during the whole 7 km ride on the simulator track (typically about 6 minutes). The voice of the test person was recorded and evaluated after the session.

E. Tactile Detection Task, TDT

For the TDT, a self-made device with an Arduino Uno was used. The device was set to a vibration stimulus randomly distributed from 3 to 5 seconds. The test person has to react within 2 seconds after stimulus onset, else it is a *miss*. The stimulus lasts one second or until the subject reacts. If the reaction time (RT) is lower than 200ms it is canceled a *cheat*. RTs between 0.2s and 2s are *hits*. The metric *hit rate* is the number of hits divided by the number of stimuli (see [18], [19]). As proposed by [19], a data set must have a hit rate of at least 70% to be included for data analysis. The device was programmed internally to react interrupt based on subject's reaction. The standard Arduino Uno clock resolution of 4 microseconds was estimated to be enough to measure milliseconds.

The vibration stimulus was applied via a vibration motor from an old mobile phone (Alcatel One Touch Easy 302) at an open clamping voltage of 4.2V (2.8V under load) with 22mA. The motor was attached to the right wrist with a flexible wristband (figure 2), vibrating at about 125 Hz. For the reaction, a micro switch was glued to the back side of the

Figure 2. Setup of the devices in the driving simulator

steering wheel at the 10 o'clock position (the track included no sharp curves or overtaking).

F. Instruction

The participants were instructed to give the driving task the highest priority. Their second priority was the detection task. Finally, they were to concentrate on using the HMIs or the *COTA*.

G. Participants

Twenty-two test subjects took part. One quit due to simulator sickness, so the data set is not regarded. The data for one test person revealed that she (healthy young female) probably did not use the vibration stimulus onset for the reaction in some experiments, else the automatic switch-off was interpreted as stimulus signal, unnoticed by the experimenter. Their data set was ignored for the TDT calculations.

The age of the test subject was from 20 to 32 years (M=25.1, SD=3.1). Two females took part. All of the test persons had normal or corrected to normal visual acuity (43% used corrective lenses during the experiment). Three had a color perception deficiency. The average annual milage was 11,486 km (SD 11,825). Ninety percent had driven an automatic car (like the driving simulation car) before. Previous experience with a driving simulator was reported by 76%. 19% of the test persons took part in the first experiment for a traffic light assistant at the institute [2], [4].

IV. Results

A. Customization

- Parameters: *[carpet on/off]* & *[carpet distance X meters ./. 100%]* All of the test subjects enabled the speed recommendation. Nineteen of the twenty-one subjects wanted to be informed about the carpet whenever possible. One person specified 300 meters in front of the traffic light and another person 1000 meters before.
- Parameter *[arrival on red on/off]*: 57% preferred to be informed that they would probably arrive at red, instead of a blank screen.
- Parameter *[speed alert on/off]* & *[speed tolerance X km/h]*: 71% enabled the speed alert at an average speed tolerance of 17.5 km/h above the allowed speed limit.
- Parameter *[Heuer on/off]*: 62% of the subjects selected the Heuer traffic light option for their individual HMI.
- Parameter *[distance bar on/off]*: 62% wanted the distance bar in their customized HMI.
- Parameter *[residual red on/off]*: 91% wanted to be informed about the remaining red light time while waiting at the traffic light.
- Parameter *[intersection suppression on/off]* & *intersection radius X meters*: Three person selected the suppression of system information at intersections at radiuses of 100, 150 and 200 meters.

B. Tactile Detection Task, TDT

The TDT results were only included for a person in an experimental section if the hit rate was above 70% (see [19]). For *COTA*, two data sets do not meet this quality criterion, and one data set for *HMI complex*. A statistical test reported no significance between the reaction times (Figure 3) of *Baseline*, *HMI individual*, *HMI complex* and *COTA*). In the *COTA* condition, the *COTA*-software pronounced on average 26.5 challenges. 99% of these were answered correctly. This shows that the subjects were all engaged in doing the cognitive task.

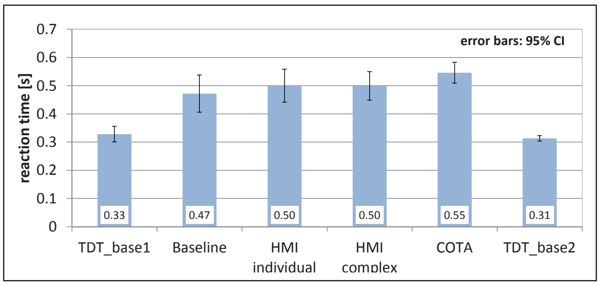


Figure 3. TDT reaction times

Figure 4 shows the hit rates under different conditions (data sets with a hit rate lower than 70% were also included). A statistical test reported no significance between *Baseline*, *HMI individual*, *HMI complex* and *COTA*).

C. System Usability Scale, SUS & AttrakDiff2

The system usability scale (SUS) reported:

- *HMI complex* 72.6 (SD 12.6)
- *HMI individual* 73.9 (SD 12.9)

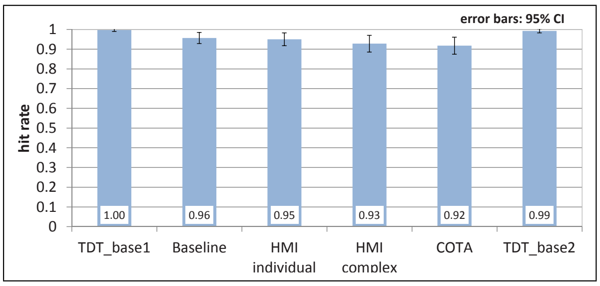


Figure 4. TDT hit rates

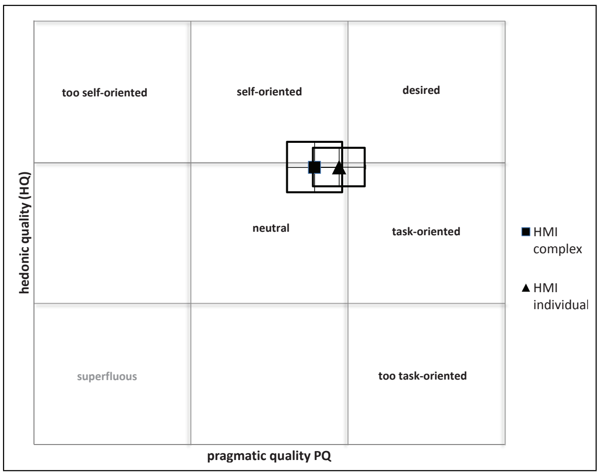


Figure 5. AttrakDiff2 portfolio diagram with confidence rectangles

The attractiveness dimension (scale from 1 to 7) of the AttrakDiff2 reported

- *HMI complex* 4.7 (SD 0.6)
- *HMI individual* 4.9 (SD 0.9)

The other dimensions of the AttrkDiff2 are shown in a portfolio diagram (Figure 5).

D. Gaze Behavior

For the first analysis the gazes were exported from D-Lab to Matlab. In Table I and Table II, the gazes of all test subjects are handled as a whole. Thus, a reported 85th percentile value (p85) or mean value (avg) is the p85 or mean value of the number (N) of gazes. The table only includes values for gazes, while the smart phone showed a speed recommendation (*velocity carpet*) and all other conditions are initially neglected (arrival on red, residual red, etc.). The average frequency (fq) is the entire duration of the *velocity carpet* condition divided by the number of glances by all subjects. In the *baseline* run, only 3 gazes toward the AOI *smart phone* are recorded, in *COTA* 8. So no avg and p85 values are reported. Accordingly, the eyes-off-the road incidents are mainly *speedometer* gazes. For the *eyes-off-the-road* values, the gaze durations not directed toward the windscreen are evaluated. In the conditions *baseline* and *COTA*, the display of the smart phone is blanked out (black), but in order to get segments for comparison, the smart phone nevertheless reports to the eye tracking system whether a *velocity carpet* would have been shown. The percentage of time (%) is calculated: frequency (fq) multiplied by mean glance (avg)

TABLE I. GAZES TOWARD THE SMART PHONE

	N	fq [Hz]	avg[s]	p85[s]	%
Baseline	3	0.001	n.a.	n.a.	n.a.
HMI individual	1072	0.275	0.64	0.88	17.5%
HMI complex	802	0.259	0.65	0.91	16.9%
COTA	8	0.004	n.a.	n.a.	n.a.

TABLE II. EYES OFF THE ROAD

	N	fq [Hz]	avg [s]	p85 [s]	%
Baseline	410	0.272	0.69	0.92	18.7%
HMI individual	1580	0.406	0.85	1.20	34.3%
HMI complex	1240	0.400	0.82	1.20	33.0%
COTA	516	0.256	0.68	0.88	17.3%

For the *HMI complex* run, D-Lab reported a mean gaze duration of 0.61s and a gaze frequency of 0.20Hz, if the smart phone showed the *arrival on red* screen. In addition, for *HMI complex*, a mean glance duration of 0.38s can be derived for the speed alert screen from D-Lab (with an average frequency of 0.24Hz). In *HMI complex* (zero speed level tolerance for speed alert system), the eye tracking system logged a mean total duration for speed alerts of 80s per run (mean duration of run: 348s). So the test subjects drove about 23% of the time with a 'nag screen'.

E. Driving Behavior

The time that each driver drove above the speed limit was set in relation to the time the car was moving faster than 5km/h. The average of these percentage values for each person's run can be seen in Figure 6.

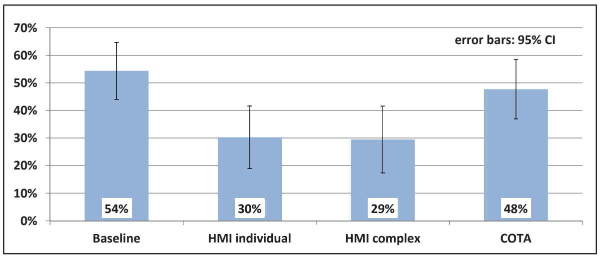


Figure 6. Average percentage of time (while car moving >5km/h) above speed limit

For Figure 7, the excess speed beyond the limit was treated as root mean square value (RMS). Values below the speed limit were treated as zero. The average speeding RMS of the test subjects' single runs is reported.

 ISBN: 978-1-61208-325-4

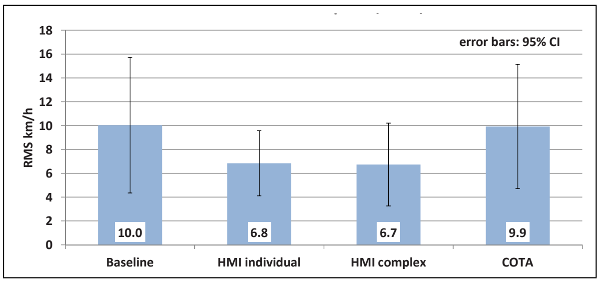


Figure 7. Average RMS value above allowed speed (km/h)

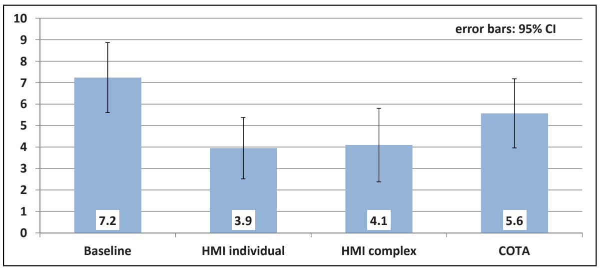


Figure 8. Average number of speed violations

The person's speed violation counter was triggered if the speed exceeded the allowed value by >15km/h. After a detection, the next re-triggering for a rising edge above 15km/h is inhibited for 10s. The average value for the single test person's run can be found in Figure 8.

V. Discussion

The customization shows that the test persons wanted the speed recommendation, not at a fixed distance, but whenever possible. More than half of the test subjects also wanted to be informed whether they would arrive on red. About two-thirds enabled the speed alert. This acceptance is probably closely coupled with the high average tolerance speed of 17.5 km/h. If one take into consideration, that the square of speed is included in kinetic energy (crash), the speed estimation in reality will likely come from a 1Hz GPS receiver, and German law enforcement on rural roads will most likely use a tolerance level of about 12-14 km/h; the value of 17.5 km/h is too high. A tolerance level of 10 km/h, which might work in reality, was accepted by one-third of the test subjects and was implemented for later tests on real roads. The Heuer traffic light sign and the distance bar had a popularity level of over 50%. A later expert review revealed that there would be too many moving and animated screen objects. It was decided to drop the distance bar in order to get a clearer presentation. The remaining red light time while waiting was enabled by over 90%. That is positive indicator for acceptance. The option to suppress system information in intersection areas was not used by 85%. That is a clear sign of rejection.

The TDT results from *TDT_base1* and *TDT_base2* are very close. The reactions from *TDT_base2* are a little bit faster. So, learning effects and fatigue did not play an important role or cancel each other out. The standard deviation (and thus the related confidence interval) for *TDT_base2* gets very small. This indicates that at the end of the experiment there is not much intersubject difference. Reaction times and hit rates show a plausible order: *Baseline* imposes the lowest cognitive load. *HMI individual* and *HMI complex* are about equal, and a little bit more mentally demanding than *Baseline*. The cognitive most demanding condition appears to be *COTA*.

The SUS values of *HMI complex* (72.6) and *HMI individual* (73.9) are very close. The individual customization is not reflected by a high gain in the subjective usability scale rating. According to [20], both of the SUS values reported here, can be associated with the adjective *good*. In the first simulator study without the TDT [2], a nearly equal value of 75.3 was found. The slight drop may be is influenced by the TDT. The attractiveness dimension of the AttrakDiff2 (from 1 to 7) also reports only a minor change between *HMI complex* (4.7) and *HMI individual* (4.9). The value from the first experiment [2] was 5.0. The portfolio diagram shows that the hedonic quality for *HMI complex* and *HMI individual* are at the same level, but the pragmatic quality is rated a little bit higher for *HMI individual*.

Often, the 85th percentile value for the glance duration is calculated using the 85th percentile of the mean values for the test subjects. The maybe more conservative way, obtaining the 85th percentile value of all gazes, reports durations that are still in line with guidelines for single glances. Results of a later real road experiment show that the gaze frequency drops in real traffic [3].

The driving behavior shows potential safety benefits in different dimensions. The total speeding time is reduced, as well as the RMS velocity of violations and the general level of illicit behavior. Similar results were found in the first simulator experiment [4], where the smart phone traffic light assistant reduced the speeding percentage of time from 60% to 25%, the speeding RMS from 12.1km/h to 6.9km/h and the number of speed violations from 6.8 to 3.6. It is interesting to note that in the first experiment no speed alert system was implemented. Thus, the main effects for speed reductions should come from the traffic light assistance and the compliant behavior of the drivers. The incorporation of a screen-filling speed alert on the phone nevertheless has the advantage that it can be easily detected and read quickly. Another benefit is that this ‘nag screen’ makes the system resistant to potential misuse (over speeding). It is also interesting to note that [8] reported a reduction of speeding and crashes on a real road segment after installation of a road side traffic light assistant.

VI. Conclusion

The improvements in the subjective usability ratings for a customized traffic light assistant on a smart phone are so small that the later system was not made highly adaptable for the user. The later system included items and presets that are accepted by most of the users, combined with safety-related ideas. The TDT values show a plausible order for the cognitive load of the experimental conditions. The traffic light assistant seems to impose only a minor additional demand on normal driving. And less cognitive load than a simple auditive number task. A speed alert would be accepted by many people, but with a relative high speed tolerance level. It will be a challenge for a safe real-life system to find the right trade-off. From the simulator results, there are no safety-related issues that inhibited tests of the carefully designed system on the road.

VII. Outlook

In the next stage of the KOLIBRI project, the traffic lights in the test fields were switched to a coordinated fixed time scheme (green wave). This was used to test the traffic light assistance on the road. The switching times were pre-determined. Thus, a prototype on a smart phone for experiments could be set up in order to test TDT, gaze behavior and driving in subject tests under real conditions.

Acknowledgments

This work was supported by Bayerische Forschungsstiftung (Bavarian Research Foundation). The authors would like to thank the project partners TRANSVER, BMW Group and the Center for Traffic Management of the Bavarian Road Administration.

References

[1] Commission recommendation of 22 December 2006 on safe and efficient in-vehicle information and communication systems: update of the European Statement of Principles on human machine interface (2007/78/EC).

[2] M. Krause and K. Bengler, "Traffic Light Assistant - Evaluation of Information Presentation". In G. Salvendy & W. Karwowski (Eds.), Advances in Human Factors and Ergonomics 2012. Proceedings of the 4th Ahfe Conference 21-25 July 2012, CRC Press, pp. 6786-6795.

[3] M. Krause, V. Knott, and K. Bengler, "Traffic Light Assistant Can Take My Eyes Off of You". Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference (submitted).

[4] M. Krause and K. Bengler, "Traffic Light Assistant - Driven in a Simulator." In R. Toledo-Moreo, L. M. Bergasa, & M. . Sotelo (Eds.), Proceedings of the 2012 International IEEE Intelligent Vehicles Symposium Workshops.

[5] C. Voy, W. Zimdahl, W. Mainka, and F. Stock, 1983, German Patent DE3126481C2.

[6] G. Hoffmann, "Up-to-the-minute information as we drive - how it can help road users and traffic management". Transport Reviews, vol. 11, no. 1, pp. 41-61, 1991.

[7] R.S. Trayford, B.W. Doughty, and J.W. van der Touw, "Fuel economy investigation of dynamic advisory speeds from an experiment in arterial traffic". Transportation Research Part A: General, vol. 18, no. 5-6, 1984, pp. 415-419.

[8] R. Trayford and T. Crowle, "The ADVISE traffic information display system". In Vehicle Navigation and Information Systems Conference 1989. Conference Record, 1989, pp. 105-112.

[9] X H. Chen, Zhou Huaguo, and P. Hsu, "What Do We Know About Signal Countdown Timers?". ITE Journal on the web, vol. July 2009, pp. 72-76.

[10] J. Pauwelussen, M. Hoedemaeker, and K. Duivenvoorden, CVIS Cooperative Vehicle-Infrastructure Systems: Deliverable D.DEPN.4.1b Assess user acceptance by small scale driving simulator research. 2008.

[11] S. Thoma, Mensch-Maschine-Interaktionskonzepte für Fahrerassistenzsysteme im Kreuzungsbereich". Dissertation, TU München, 2010.

[12] R. Braun, F. Busch, C. Kemper, R. Hildebrandt, F. Weichemeier, C. Menig, I. Paulus, and R. Prelein Lehle, "TRAVOLUTION - Netzweite Optimierung der Lichtsignalsteuerung und LSA-Fahrzeug-Kommunikation, Straenverkehrstechnik". vol. 2009, no. 6, pp. 365-374.

[13] simTD, "Sichere Integrierte Mobilität - Testfeld Deutschland". German C2X project 2008-2013, http://www.simtd.de [retrieved Dec. 2013].

[14] aktiv, "Adaptive und kooperative Technologien für den intelligenten Verkehr" German Research Initiative 2006-2010, http://www.aktiv-online.org page [retrieved Dec. 20123].

[15] R. Hoyer, "Verkehrliche Potenziale des vorausschauenden Fahrens an kooperativen Lichtsignalanlagen". 5. Tagung Fahrerassistenz München, 2012

[16] E. Koukoumidis, L.-S. Peh, and M. R. Martonosi, "SignalGuru: leveraging mobile phones for collaborative traffic signal schedule advisory". In Proceedings of the 9th international conference on Mobile systems, applications, and services, New York, NY, USA: ACM, 2011, pp. 127-140.

[17] C. Olaverri Monreal, P. Gomes, M. Krüger Silvéria, and M. Ferreira, "In-Vehicle Virtual Traffic Lights: a Graphical User Interface". 7th Iberian Conference on Information Systems and Technologies, CISTI'2012, Madrid, Spain, pp. 1-6.

[18] J. Engström, "The Tactile Detection Task as a method for assessing drivers cognitive load". In G. Rupp (Ed.) Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety. Warrendale, PA: SAE International. 2010, pp. 90-103.

[19] N. Merat, E. Johansson, J. Engström, E. Chin, F. Nathan, and T. Victor "Specification of a secondary task to be used in safety assessment of IVIS". AIDE Deliverable 2.2.3 IST-1-507674-IP. 2006.

[20] A. Bangor, P. T. Kortum, and J. T. Miller, "An Empirical Evaluation of the System Usability Scale". International Journal of Human-Computer Interaction, vol. 24, no. 6, 2008, pp. 574-594.

[21] J. Brooke, "SUS: A quick and dirty usability scale". In P. W. Jordan, B. Weerdmeester, A. Thomas, & I. L. Mclelland (Eds.), Usability evaluation in industry, pp. 189-194 London: Taylor and Francis, 1996.

[22] M. Hassenzahl, M. Burmester, and F. Koller. "AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität". In: J. Ziegler & G. Szwillus (Hrsg.), Mensch & Computer 2003. Interaktion in Bewegung, pp. 187-196, Stuttgart, Leipzig: B.G. Teubner.

Evaluation of Window Interface in Remote Cooperative Work Involving Pointing Gestures

Ryo Ishii, Kazuhiro Otsuka
NTT Communication Science Laboratories,
NTT Corporation
Kanagawa, Japan 243–0198
Email: {ishii.ryo, otsuka.kazuhiro}@lab.ntt.co.jp

Shiro Ozawa, Harumi Kawamura and Akira Kojima
NTT Media Intelligence Laboratories,
NTT Corporation
Kanagawa, Japan 239–0847
Email: {ozawa.shiro, kawamura.harumi, kojima.akira}@lab.ntt.co.jp

Yukiko I. Nakano
Faculty of Science and Technology,
Seikei University
Tokyo, Japan 180–8633
Email: y.nakano@st.seikei.ac.jp

***Abstract*—Gazes and pointing gestures are important in performing collaborative work involving instructions with shared objects. However, in general video conferencing systems, the geometrical consistency of size and positional relationships of remote spaces are not displayed correctly on the display screen. This inhibits the transmissions of gazes and pointing gestures vis-a-vis shared objects. It is thus important to demonstrate how gazes and gestures can be smoothly transmitted by video and develop an advanced system that can do it. We previously proposed a "MoPaCo" window interface system that can reproduce a communication partner's space within a display as if the display were a glass window to achieve geometrical consistency between remote spaces. Experiment results demonstrated it enables users to feel the distance between themselves and their conversational partners on video is about the same as in a face-to-face situation and the partner is actually present. We also consider MoPaCo can generate video images that smoothly transmit gazes and pointing gestures; this paper describes experimental tests of the system's effectiveness in doing so. Results suggest MoPaCo allows users to accurately identify target objects as they could under face-to-face conditions through an actual glass window. Results of experiments on conversation quality show MoPaCo facilitates smooth conversation and communication among users and strengthens their memories of the conversations, suggesting the users actively engage in conversation and the system makes a strong impression on them.**

Keywords-Remote cooperative work; full gaze awareness; pointing gesture; window interface.

I. INTRODUCTION

Our objective is to achieve an advanced media space providing a seamless connection between two remote spaces. This will enable users to work closely together while sharing their respective spaces, discuss things, such as furniture layouts, and smoothly perform collaborative work involving the following of operating instructions. As an example of this, we simulate a situation where users in two seamlessly connected remote places discuss a certain burden and an appropriate place to put it before transferring it from their space to a remote space. In such a situation, the media space is expected to enable smooth transmission of nonverbal behavior such as gazes and pointing gestures (hereafter "gestures"). Nonverbal behavior is known to play an important role in ensuring smooth performance of collaborative work and instruction work [1], [2], [3]. However, since video images are displayed as-is in general videoconferencing systems, the geometrical consistency of size and positional relationships of the remote spaces are not displayed correctly. Thus, gaze and gesture directions cannot be correctly transmitted [4]. A major topic in human-computer interaction research has been the need for an advanced system and method giving users video images that look like face-to-face situations and allow smooth transmission of their nonverbal behavior. However, to date, no such method or system has been developed for an environment where two remote spaces are connected and actual objects are shared in them.

We previously proposed a window interface system called "MoPaCo" that reproduces a communication partner's space within a display as if the display were a glass window to achieve geometrical consistency between two remote spaces [5]. Since MoPaCo imparts motion parallax that adjusts to a user's viewpoint position, users can feel as if the remote spaces are connected smoothly as if separated only by a glass window. Experiment results demonstrated the users feel the distance between themselves and their conversational partners on video is about the same as in a face-to-face situation where the partner is actually present [6]. Since MoPaCo achieves geometrical consistency between two remote spaces, it is considered to have excellent potential for enabling smooth transmission of gazes and gestures. However, its effectiveness in doing so has never been tested. If this could be demonstrated, it would demonstrate that achieving video images connecting two remote spaces seamlessly as if they were separated merely by a glass window would be effective in transmitting gazes and gestures. This knowledge would make a significant contribution as a guide for designing new remote collaborative systems.

This paper describes experiments conducted to determine whether MoPaCo accurately transmits gazes at and gestures made to shared objects. It also describes evaluation experiments performed involving remote collaborative work to determine whether correct gaze transmission positively affected communication smoothness. The results indicate the system allows gazes and gestures to be transmitted in a similar manner

 ISBN: 978-1-61208-325-4

as in face-to-face conditions. They suggest MoPaCo users could refer to target objects smoothly, as if speaking face-to-face through a glass window, and conversation partners could predict the next target to be explained. Subjective assessments indicate MoPaCo encourages natural conversation and communication, facilitates conversation smoothness, and strengthens users' memories of conversations. This demonstrates the system contributes to improved conversation quality.

In the rest of this paper, Section 2 reviews related work and highlights of the paper. Section 3 presents details of the MoPaCo system. Sections 4 and 5 describe the evaluation of the system's gaze and gesture transmission and the evaluation experiments conducted involving remote collaborative work. Section 6 discusses the evaluation results in detail and Section 7 concludes the paper with a summary.

II. RELATED WORK

A. Importance of gazes and pointing gestures

Nonverbal behavior is known to play an important role in social psychology for performing collaborative work and instruction work smoothly. When conversation participants share the same physical environment and their tasks require complex reference to and joint manipulation of physical objects, the participants frequently observe a shared object most of the time instead of paying direct attention to their partner [1], [2], [7]. In such situations, establishing joint attention by paying attention to the shared object signals the listener's engagement in the conversation, and functions as evidence for comprehension in conversation grounding [8]. For example, if the listener asks for directions while observing a map, the listener's behavior in directing his or her gaze at the map to indicate sharing of the map information gives effective nonverbal feedback serving as evidence of comprehension. Suzuki et al. analyzed the relationship between gaze behavior and task completion, demonstrating nonverbal information such as gazes and gestures governs the success of a task [3].

In indication work, when referring to a target object in an indication, projectability, i.e., the predictability of which object a partner is observing and what he or she will explain or do from the direction of a partner's body or gaze, is shown to be important in making reference to objects easy [9]. In connection with this finding, Goodwin analyzed nonverbal behavior under face-to-face conditions, in which a speaker indicates a target object to a listener [10]. First, the listener appropriately adjusts the direction of his or her body so as to share a mutual gaze at the object with the speaker. This indicates the listener is actively listening to the speaker. Conversely, when the speaker gives the indication, he or she changes position so both the object and listener are visible. The listener then comprehends the speaker's target of interest and directs attention at the next object to be indicated. In this way, when the speaker refers to a target object, the listener can smoothly identify it.

B. Gazes and gestures in communication systems

Video expression enabling transmission of gazes and gestures has been a major challenge in human-computer interaction research. Here, a person's awareness of the conversational partner's gaze is defined as gaze awareness. Gale and Monk divided gaze awareness into three levels, as follows [11].

- Mutual gaze awareness: A person can understand he/she is being observed by a conversational partner. This is generally known as "eye contact".
- Partial gaze awareness: A person can understand the eye direction (up, down, left, right) of the conversational partner.
- Full gaze awareness: A person can understand what object the conversational partner is observing.

This classification also applies to gestures. Many studies have focused on achieving gaze awareness in video conferencing systems. First, methods for achieving mutual gaze awareness in remote face-to-face communication have been considered. Methods have been proposed using a half mirror [12], a liquid crystal shutter [13] and a stereoscopic camera or time-of-flight camera [14] to generate frontal facial images. In addition, a widely used method has been developed in which the deviation of the face and camera positions is five degrees or less and thus eye contact is achieved [15].

Furthermore, systems have been proposed for extending multi-party conversations, i.e., HYDRA [16], Browser Magic [17], GAZE Groupware system [18], and GAZE-2 [19]. These systems enable users to understand the direction a person faces from the person's head direction; thus both partial and mutual gaze awareness are achieved. In addition, the Browser Magic system [20] enables users to understand whom the conversation partners are observing; thus, full gaze awareness is achieved assuming users at three remote sites. Furthermore, a method has been proposed to achieve full gaze awareness in many-to-many human conversations, i.e., MultiView [21], which presents parallax images in accordance with each user's viewpoint using a camera and projector for each user. However, these systems focused on who the participants observe and did not address the issue of correctly transmitting gaze behavior to objects in shared spaces. Although Clearboard [22] enables gazing at a shared display surface, it is limited to the display surface and does not achieve full gaze awareness for objects in a shared space. Therefore, insufficient study has been done on video expression techniques connecting two remote spaces in a media space smoothly and achieving full gaze awareness allowing users to understand what objects their conversational partner is observing.

In another attempt to achieve effective transmission of gazes at and gestures made towards shared objects in a remote space, the idea of having a vicarious robot stand in for the user has been proposed [23]. This robot acts as a substitute for a remote user and reflects gestures and head direction (pseudo eye direction) in real time. In a test at a surrogate robot exhibition, it was able to smoothly establish mutual gazing by directing its attention to audience members observing it and referring to objects pointed out to it. This research showed the importance of transmitting gazes and gestures in achieving smooth remote communication, but focused only on transmitting three nonverbal information factors (gazes, gestures, and body positions), using a vicarious robot device as a human substitute. However, there is a need to convey multiple, complex nonverbal information factors in addition to gazes, gestures, and body positions, e.g., facial expressions and nodding. From this viewpoint, it must be considered important to transmit all nonverbal information emanating from the person in the video to transmit nonverbal behavior in the same manner as in face-to-face situations.

In contrast to these methods, our aim is achieving geometrical consistency for the size and positional relationships of two remote spaces on a video display. We suggest the MoPaCo

 ISBN: 978-1-61208-325-4

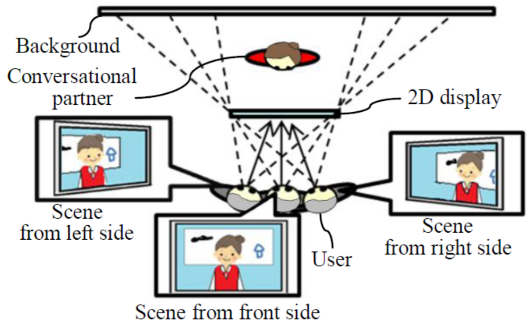


Figure 1: Concept images of video representations caused by motion parallax.

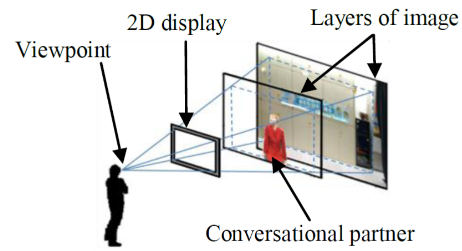


Figure 2: The person layer and background layer are projected.

system as a means of presenting images as clearly as if the spaces were merely separated by a glass window [5]. Since MoPaCo reproduces the size of the spaces and their positional relationship, it transmits body positions, gestures, and gazes naturally and correctly. We have previously performed experiments with the system demonstrating it allows users to feel the interpersonal distance between themselves and their conversational partners in a remote space so they can feel the reality of face-to-face communication and the partner's presence [6]. Since MoPaCo achieves geometrical consistency between two remote spaces, it is considered to have excellent potential for smoothly transmitting gazes and gestures.

III. WINDOW INTERFACE: MOPACO

A. System Summary

We previously proposed a real-time video communication system called Motion Parallax Communication (MoPaCo) that reproduces a communication partner's space within a display as if the display were a glass window to achieve geometrical consistency between two remote spaces [5], [6]. Figure 1 shows MoPaCo-produced motion parallax video images of a conversational partner that correspond to the viewpoint positions of different users. The display for a user some distance from the partner in the video can give the user and partner the feeling they are linked as if seeing each other through a glass window. We consider this motion parallax video representation will eliminate spatial separation, improve the conversational partner's presence, and enable the transmission of nonverbal information associated with depth by imparting depth information to video images. Presenting a motion parallax video of a partner on a 2D display corresponding to the viewpoint positions of different users requires the following process:

(I) Measuring each user's viewpoint position.

(Ⅱ) Constructing a 3D space having information on the dimensions and positional relationships of the people and the background, based on information obtained from a camera or other means.

(Ⅲ) Rendering the 3D space constructed in step (II) on a 2D display, to correspond to each user's viewpoint position obtained in step (I).

MoPaCo implements steps (I) and (II) with a single monocular camera. This section describes the detailed process for steps (I) to (III).

B. Measuring User's Viewpoint Position

We proposed using a single monocular camera to detect each user's viewpoint. Before calculating the 3D position from parts information (coordinate position) of each face in the 2D image, the system performs preprocessing by measuring the eye separation distance of each user. It then acquires the distance of each user from the camera, using the depth from focus function used for achieving focus in ordinary cameras. The lens distortion of the image was eliminate by Zhang's lens distribution correction method [26] before this process. During this process, template matching is performed on the image captured from the camera to measure the positions of both eyes (2D coordinates within the image) and the orientation of the head. The system calculates the eye separation distance of each user from the user-to-camera distance, the information measured from the image, and the camera's angle of view and resolution. With this information, real-time capture starts and the system obtains the positions of both eyes (2D coordinates within the image) and the orientation of the head from the captured image, and calculates the viewpoint position z of that user from the camera from there at that time. The x- and y-coordinates are calculated from the 2D coordinates within the image and the image's pixel pitch.

C. Construction of 3D Space

We proposed constructing 3D information for an image captured from a single camera by performing background difference processing using background information acquired beforehand (Images of several seconds were captured for background information), maintaining the 2D plane and dividing it into personal and background areas, and creating a multi-layer structure with those areas arranged as layers in accordance with their depth-wise positions (see Figure 2). Using 2D images ensures a high-resolution display; furthermore, subjecting only the background difference to image processing lowers processing costs and enables real-time processing. The system generates a "person layer" showing a full size image of a person and a "background layer" showing a full size image of the background. These layers have a distance relationship from the camera. The distance information measures for the background layer are calculated beforehand using the depth-from-focus method of the camera's auto focus function, when the background difference image is acquired. For the person layer, the user viewpoint position is used. These distances become the information about the distance from the camera to the person layer and the background layer, respectively. The system then uses (1) to calculate the full size (width w_i × height h_i) of each layer i from the thus-acquired distance information d_i and the camera's angle of view (width θ_x, height θ_y). This procedure configures a 3D space having full size and position information.

$$w_i = 2 * d_i * tan(\theta_x/2), h_i = 2 * d_i * tan(\theta_y/2) \quad (1)$$

 ISBN: 978-1-61208-325-4

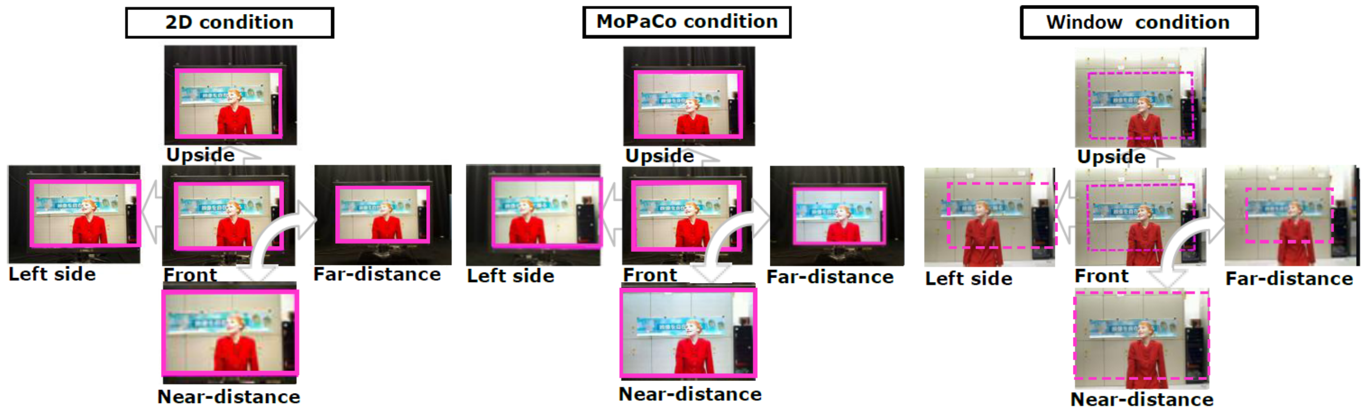


Figure 3: Scenes for 2D, MoPaCo, and Window conditions.

D. Rendering 3D Space on User Viewpoint Basis

As shown in Figure 2, the person layer and background layer generated by the 3D spatial information module are projected in perspective to match the user's viewpoint position, using the 2D display as a projection surface. Thus, motion parallax video is implemented.

E. Implementation

Using the above-described methods, we implemented the MoPaCo system, which enables real-time bidirectional viewing of motion parallax video. The development environment was a camera with HD size resolution (1920 × 1080), a computer with Intel Core i7 Extreme 980X as the CPU and 12 GB of memory, and a NVIDA GeForce GTX480 graphics board. Table I shows the implementation results; "lag from viewpoint movement" is the time from the user's viewpoint position moving to the time motion parallax appears in the video; "lag of camera image" is the time until the captured video appears.

Figure 3 shows scenes used in experiments conducted to enable users to evaluate the MoPaCo system. They show differences in the visibility of a conversational partner from the same position under 2D, window, and MoPaCo conditions. Under window conditions, users can observe the conversational partner through an actual glass window. Five view positions were used, i.e., the front, left, and upper sides of the display, and the near distance to and far distance from the display. In comparing the scenes under the window and 2D conditions, since there was no parallax in the video under the 2D conditions even if the user's head moved, the human dimensions and positional relationships did not match. Under the MoPaCo condition, in contrast, the dimensions and positional relationship between the person and background were reproduced in the video.

IV. Evaluation of Accuracy of Gaze and Pointing Gesture Transmission to Objects

A. Experimental Method

We conducted experiments to determine whether the MoPaCo system correctly transmitted gazes and gestures, and to compare and verify transmission accuracy when it was directed through the window and the actual 2D video as general experimental conditions, in addition to MoPaCo conditions.

- 2D condition: observing the conversational partner in an image taken with a camera directly on a 2D display. This condition is for the use of a classic 2D video conferencing system. In this case, the user's viewpoint position is where the image is displayed at a position when the user is sitting straight in the chair.
- MoPaCo conditions: observing the conversational partner with MoPaCo.
- Window conditions: observing the conversational partner through a glass window.

TABLE I: Performance of MoPaCo system.

	Frame rate	Response
Lag of camera image	30 fps	260 ms
Lag from viewpoint movement	30 fps	300 ms

B. Experiment Results

In the experimental setup, the subject was seated on a chair 80, 150, or 230 cm from a partition with a glass window installed between the subject and his/her conversational partner (Figure 4) and was able to observe the partner's space through the glass window. Since the glass window size (46 cm high × 80 cm wide) was less than the display size, the subjects could not see the display edges. A camera was installed immediately above the glass window so as to match the participant's gaze [15].

Rectangular 50 × 50 panels for use as indication targets were placed in a 3 × 18 panel arrangement on a wall 200 cm behind the participant (panel rows were labeled from A to C vertically from the top; columns were labeled 1 to 18 from the left). Four panels were chosen as indication targets:

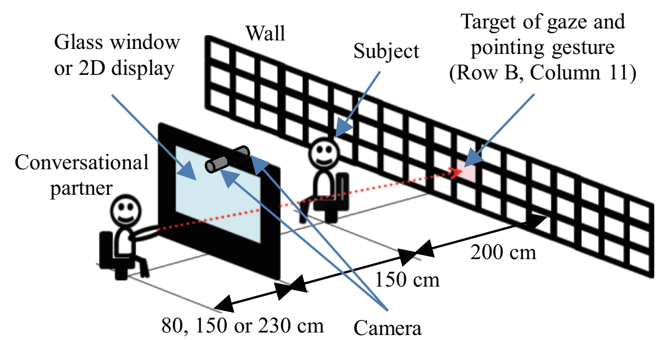


Figure 4: Experimental setting for measuring gaze and pointing gesture transmission accuracy.

 ISBN: 978-1-61208-325-4

TABLE II: Pointing gesture accuracy rate and Turkey-Kramer multiple comparisons.

Conditions		Accuracy rate (%)			Multiple comparison		
Targets	Partner's position	2D	MoPaCo	Window	2D vs MoPaCo	2D vs Window	MoPaCo vs Window
Row B, Column 4	80 cm	3.0	21.2	24.2	*	*	n.s.
Row B, Column 4	150 cm	0	27.3	24.2	**	**	n.s.
Row B, Column 4	230 cm	0	13.3	15.2	n.s.	n.s.	n.s.
Row B, Column 8	80 cm	3.0	30.3	33.3	*	*	n.s.
Row B, Column 8	150 cm	3.0	30	27.3	**	*	n.s.
Row B, Column 8	230 cm	0	24.2	27.3	*	*	n.s.
Row B, Column 11	80 cm	3.0	33.3	33.3	*	**	n.s.
Row B, Column 11	150 cm	0	27.3	30.3	*	†	n.s.
Row B, Column 11	230 cm	0	21.2	24.2	†	*	n.s.
Row B, Column 15	80 cm	0	12.1	21.2	†	†	n.s.
Row B, Column 15	150 cm	0	9.1	9.1	n.s.	n.s.	n.s.
Row B, Column 15	230 cm	0	9.1	9.1	n.s.	†	n.s.

†: p<.10, *: p<.05, **: p<.01

B4, B8, B11, and B15. During the experiment, the participant was shown someone performing a gesture either in a video or through the window and verbally answered which object was being indicated. Three trials were performed for each condition. To minimize order effects, experiment conditions were randomly chosen from combinations of three observation conditions, three indicator positions, and the four indication targets.

Table II shows the experiment results obtained for the 11 participants (9 males and 2 females in their 20s). The table shows the average accuracy rate of participant answers regarding the indication target under each of the three experiment conditions. We performed a repeating two-way factorial analysis of variance for each of the four indication targets to determine whether the conversation partner's position or observation conditions affected the accuracy rate. This showed the conversation partner's position did not have a significant effect but the observation conditions did (B4: $F(2, 90) = 13.92, p<.01$, B8: $F(2, 90) = 10.23, p<.01$, B11: $F(2, 90) = 15.56, p<.01$, B15: $F(2, 90) = 8.14, p<.01$). Since the observation conditions had a contributing effect, multiple comparisons were performed for each of the three observation conditions using the Tukey-Kramer method. Table II shows the test results; the accuracy rates for the 2D condition were 0% in most cases but increased dramatically under the MoPaCo and window conditions, showing significant differences and trends. No significant differences were seen between the results for the MoPaCo and window conditions. This shows similar precision is obtained regardless of distance when transmitting indication actions under the MoPaCo and window conditions, i.e., MoPaCo successfully reproduces an actual window's size and location relationships.

V. EXPERIMENT IN REMOTE COLLABORATION INVOLVING POINTING GESTURES

A. Experimental Procedure

We investigated the communication smoothness MoPaCo provides using nonverbal behavior such as gazes and gestures in remote collaborative work. Specifically, we evaluated video expression achieving geometric integrity involving actual size and position relations with motion parallax adjusting to the user's viewpoint to confirm MoPaCo achieves smooth communication by smoothly transmitting gazes and gestures. As the evaluation method, we propose a hypothesis that the video expression can allow a user to recognize a shared object that the conversational partner indicates with gazes and gestures. We also propose a hypothesis that smooth transmission of nonverbal behavior such as gazes and gestures will facilitate smooth remote communication and improve users' impressions of conversations and conversation quality factors such as the user's conversational engagement. Accordingly, we evaluated the system for the smoothness and the impressions of conversations it provides.

In carrying out the evaluation, two subjects were placed in a conversational setting and tasked with choosing the furniture layout in each other's rooms. As the specific method of evaluating the smoothness with which they could identify the objects their partner indicated, we measured the time required to identify objects and the number of utterances one of the subjects had to make about the object's position before the other could positively identify the object. We expected the required time would become shorter and the number of utterances would become smaller if gazes and gestures were used to help the user identify the object. Conversation quality was assessed through 6-level subjective evaluations made using the Rickert method, with questionnaires asking questions about conversation smoothness and impressions. Subjective evaluation items are shown in Table III. In addition to subjective evaluations, we measured the participants' memory of the conversation and the furniture used as an indicator of whether they actively participated in the conversation. We consider that active participation and strong impressions of a conversation create stronger memories. Specifically, 80 pieces of furniture were shown in the questionnaire form and the subjects answered whether a given piece of furniture was in the partner's room. Then, we measured the accuracy rate of the subjects' responses.

To perform a comparative investigation between ordinary conversations, 2D video conversations, and conversations through an actual glass window, tests for this activity were conducted under these conditions.

- 2D condition: a conversation through images taken with a camera displayed as-is on a 2D display (the display and

TABLE III: Contents of subjective evaluation.

· Conversation smoothness: Did the conversation progress smoothly?
· Communication: Was communication achieved?
· Window feeling: Did you feel as though you were speaking through a window?
· Enjoyment: Did you enjoy the conversation?
· Affinity: Did you feel an affinity toward your conversation partner?

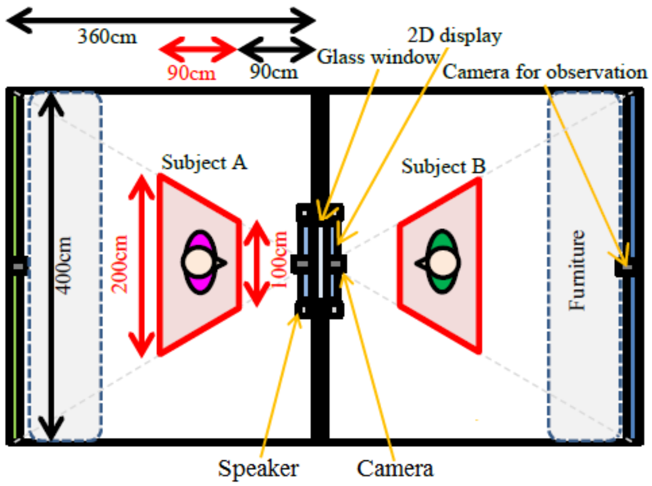


Figure 5: Top view of experimental equipment.

Figure 6: Example of arbitrarily-placed furniture.

camera angle were adjusted to include all objects to allow the participant to see the entire region of objects the indicator would refer to). This condition is for the use of a classic 2D video conferencing system.

- MoPaCo conditions: a conversation through a window image using MoPaCo.
- Window conditions: a glass window was placed between two adjacent rooms and participants conversed through it.

Figure 5 shows the experiment environment. Two participants entered adjacent rooms (360 cm × 400 cm) assigned individually to them and stood in a space in which they could move (a trapezoid 90 cm tall, 100 cm at the top, and 200 cm at the base) 90 cm away from the wall separating the two rooms. Participants were permitted to move freely within the movement space. They were not allowed to touch and move the furniture. A glass window 49.8 cm tall × 88.4 cm wide was installed 120 cm above the floor on the wall separating the rooms. Under the window conditions, conversations took place through this window. Under the 2D and MoPaCo conditions, a 40-inch 2D display (1920 × 1080 resolution) identical in size to the window was installed in front of the window. Participants communicated while watching the video. Image delay was 300 ms under both 2D and MoPaCo conditions. A camera was installed immediately above the window so as to match participant gaze [15]. Voice was collected through a microphone located in front of the display, and was output to speakers located directly beside the partner's window. Sound was delayed by 200 ms using a delay generator to ensure lip-sync under both 2D and MoPaCo conditions. Each room was arranged with 14 items (poster, table, TV, etc.) chosen randomly from a set of 84 items. Figure 6 shows an example object layout in the room.

The experiment began with the participant standing in the center of the movable space. At a signal to begin, images and voice of the participant's partner were output, and the pair conversed for ten minutes. Participants were instructed to discuss how to preferably rearrange items placed haphazardly in the two rooms. Afterwards, participants were tasked with choosing one item from their partner's room and considering where they would place it in their own room.

To minimize order effects, the three experiment conditions were used in experiments randomly. Each pair used a different set of items under each condition. After executing the experiment under each condition, the participants filled in a questionnaire concerning subjective assessments and assessments for measuring participant memory of items in the room. Sixteen participants (10 males and six females in their 20s-40s) were formed into eight pairs of friends or family members.

B. Collected Conversation Corpus

Participants' utterances, gaze behavior, and gestures were collected for analysis through the following methods.

- Utterances: voices were recorded and transcribed.
- Gaze targets: wearable Tobii Glasses [24] were used to measure the participants' gaze behavior. This allows measurements to be taken using only a pair of transparent glasses, putting little burden on the user and avoiding blocking the view of a participant's gaze direction by covering the eyes. Tobii Glasses output the gaze location in the participant's view image as a 2D coordinate plane at 30 fps. We used the annotation tool Anvil [25] to annotate gaze target objects from video images. Each room contained 14 labeled gaze target objects and one participant.
- Pointing gestures: participant actions were collected on video images and then annotated using Anvil. Gestures were defined in three steps: "preliminary action", from when the participant began moving his or her arm to perform the gesture, "during indication", when the participant pointed at the indication target object, and "returning action", when the participant finished indicating and returned his or her arm to the starting position. Gestures were annotated in these three steps.

After synchronizing these three types of data, video and annotation data were integrated into a single file of Anvil data, and conversation corpus data was created. Figure 7 shows an example; the total data comprised 24 conversations (three conditions, eight participant pairs) of five minutes each for a total of 120 minutes of corpus data.

C. Results for Target Identification Smoothness

1) Time required for identifying objects: When a participant indicated an item using demonstrative pronouns ("here","there", etc.) the name of the item, or gestures, the time the partner needed to identify the item was measured. The time started when the item was indicated and ended when the partner started gazing at it. Since 2D and MoPaCo conditions included a 300 ms image and voice delay, the starting time was set to when the indicator's voice was output from the speakers.

 ISBN: 978-1-61208-325-4

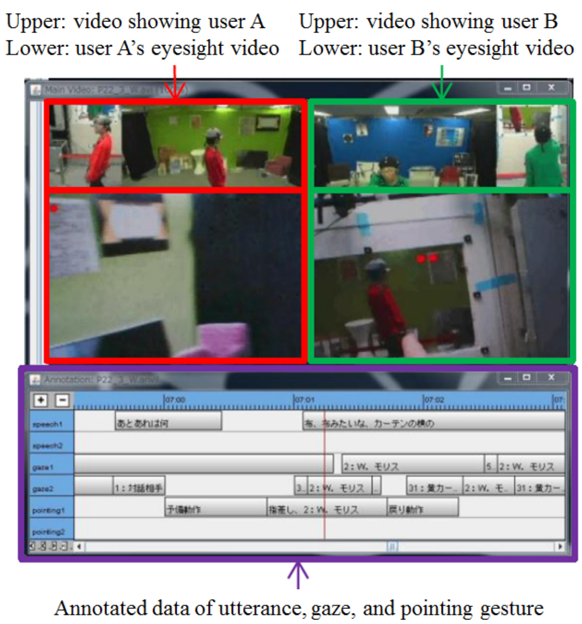


Figure 7: Corpus data in Anvil annotation tool.

TABLE IV: Result of analysis of time required for identification of indicated object.

Condition	Average time (ms)	One-way analysis of variance	Multiple comparison		
			vs 2D	vs MoPaCo	vs Window
2D	2800	*	–	**	**
MoPaCo	1100	(F(2, 447)	–	–	n.s.
Window	1500	=13.6)	–	–	–

†: p<.10, *: p<.05, **: p<.01

Table IV shows average required reference times for all conversations. The 2D conditions required the longest average time; approximately 2.8 seconds was required for 2D conditions, 1.1 seconds for MoPaCo conditions, and 1.5 seconds for window conditions. To determine whether experimental conditions made a difference in the time required for identifying target objects, we performed a one-way factorial analysis of variance. The results show a significant difference between experimental conditions ($F(2, 612) = 59.63, p<.01$). Next, we performed multiple comparisons using the Tukey-Kramer method to identify differences between pairs of conditions. These tests showed significant differences only between 2D and MoPaCo conditions ($p<.01$) and 2D and window conditions ($p<.01$). The results demonstrate 2D conditions make the identification time longer than for the MoPaCo and window conditions, and confirm MoPaCo conditions allow smooth target identification similarly to window conditions. This suggests our hypothesis was correct.

2) Number of indicator's utterances about object's position: We counted the number of utterances participants made about an object's position. Example sentences used to indicate the position included, "It's on the edge of the right-hand side of XXX (the name of another object)", "Not over there", and "It's on the opposite side". Table V shows the results obtained for the average number of utterances about an object to be identified. Under the 2D conditions the number (0.27) was highest; it was 0.06 under the MoPaCo conditions and 0.09 under the window conditions. We performed one-way factorial analysis of variance to determine whether the experimental conditions affected the differences found in the number of utterances made in indicating an object's position. The results showed there was a significant difference due to the conditions (F (2,447) = 8.38, p <.01).

TABLE V: Result of analysis of number of instructor's utterances about object's position.

Condition	Average number (per second)	One-way analysis of variance	Multiple comparison		
			vs 2D	vs MoPaCo	vs Window
2D	0.27	*	–	**	*
MoPaCo	0.06	(F(2, 447)	–	–	n.s.
Window	0.09	=8.38)	–	–	–

†: p<.10, *: p<.05, **: p<.01

TABLE VI: Result of memory of furniture in partner's room.

Condition	Accuracy rate (%)	ANOVA	Multiple comparison		
			vs 2D	vs MoPaCo	vs Window
2D	86.9	*	–	*	†
MoPaCo	94.8	(F(2,45)	–	–	n.s.
Window	93.5	=4.56)	–	–	–

†: p<.10, *: p<.05, **: p<.01

Next, multiple comparisons using the Tukey-Kramer method were performed to confirm the differences between pairs of individual criteria. Results showed significant differences between the 2D and MoPaCo conditions (p <.01) and between the 2D and window conditions (p <.05), but none between the MoPaCo and window conditions. They show subjects make more utterances to indicate an object's position under the 2D conditions than under the window and MoPaCo conditions. They also show the MoPaCo conditions enable users to identify objects with the same small number of utterances as for the window conditions. This suggests our hypothesis was correct.

D. Results for Conversation Quality

1) Memory of furniture in partner's room: We calculated the accuracy rates obtained in a memory test the subjects took regarding the furniture in their partner's room. Table VI shows the average accuracy rate for all 16 subjects' answers. We performed one-way factorial analysis of variance to verify whether the experimental conditions affected the differences found in the rate. The results showed the conditions produced significant differences (F (2,45) = 4.56, p <.05).

Next, multiple comparisons using the Tukey-Kramer method were performed to confirm the differences between pairs of individual criteria. Results showed significant differences between the 2D and MoPaCo conditions (p <.01) and between the 2D and window conditions (.05 <p <.10), but none between the MoPaCo and window conditions. They show the accuracy rate of memory about the furniture in the partner's room is lower under the 2D conditions than under the window and MoPaCo conditions. They also show the MoPaCo conditions enable users to remember conversations as well as they can under the window conditions. This suggests our hypothesis was correct.

2) Subjective evaluation results: Table VII shows the average values for participants' subjective evaluations. We performed one-way factorial analysis of variance for each of five items to determine whether experimental conditions affected the values. Since an effect of experimental conditions on the evaluation values was shown, multiple comparisons using the

TABLE VII: Subjective Evaluation Results.

Items of subjective evaluation	Average of subjective score			ANOVA	Multiple comparison		
	2D	MoPaCo	Window		2D vs MoPaCo	2D vs Window	MoPaCo vs Window
Conversation smoothness	3.0	4.0	4.3	** (F(2, 45)=11.27)	**	**	n.s.
Communication	3.6	4.5	4.6	* (F(2, 45)=3.64)	†	†	n.s.
Window feeling	3.0	4.0	4.4	* (F(2, 45)=3.42)	†	†	n.s.
Enjoyment	3.5	4.5	4.3	* (F(2, 45)=4.15)	*	†	n.s.
Affinity	3.0	4.1	4.0	** (F(2, 45)=7.25)	**	**	n.s.

†: $p<.10$, *: $p<.05$, **: $p<.01$

Tukey method were performed for each condition. Table VII shows the test results; significant differences and trends were found for each item between 2D conditions and MoPaCo and window conditions, but none between MoPaCo and window conditions. This suggests “conversation smoothness”, “communication”, “window feeling”, “enjoyment” and “affinity” were all higher under MoPaCo and window conditions than under 2D conditions, but no significant differences were found between them under MoPaCo and window conditions.

Next, we demonstrate whether the results obtained for smooth transmission of identification and improved memory about communication content have a major effect on improving communication smoothness. We evaluated the correlation between the subjective score results for items relevant to conversation smoothness and (a) the time required to identify an indicated object, (b) the number of utterances indicating the object’s position, and (c) the accuracy rate of memory about furniture. The correlation coefficient between the subjective values for conversation smoothness items and the required time was a negative correlation, -0.45. The coefficient between the values and the average number of utterances indicating the position was a low negative correlation, -0.22. This shows the differences in smoothness in identifying the object possibly affected the users’ introspection regarding the conversation smoothness. Finally, the coefficient between the subjective values for conversation smoothness and the accuracy rate of memory about furniture was a positive correlation, 0.31. This shows the differences in memories of the furniture possibly affected the users’ introspection regarding the conversation smoothness.

VI. Discussion

Evaluations of communication precision of indication actions showed that indication actions performed through MoPaCo were similarly precise to those performed through an actual glass window, regardless of the distance between the indicator and the display. We therefore consider MoPaCo successfully reproduced similar sizes and positional relationships seen in an actual glass window. While the difference was insignificant, the average accuracy rate was 23.0% under the window conditions and 21.4% under the MoPaCo conditions, i.e., the former was slightly higher. We consider this is because MoPaCo displays people as a flat layer, and thus even when users change their viewpoint their partner’s arm direction does not actually change. For example, if one participant stretches his or her arm toward another, the latter should be able to see the former’s arm stretching to the left when he or she moves to the right. In MoPaCo, the arm will still be shown stretching straight ahead. Post-experiment interviews with participants showed some of them detected a change in the direction of their partner’s arm as they moved through parallax, even though the direction did not actually change. We consider this is an illusion caused by parallax in the background. This suggests the possibility that since arm movements are slight when the user does not move much from the front of the display, even if the person is shown as a flat layer, this does not greatly affect the precision of indication actions. This leads us to consider that using MoPaCo to perform collaborative work while sharing the spaces and items in two locations allows work to progress smoothly through the natural use of indicative actions.

We conclude the required times for referring to an object indicated in a partner’s space were the same for the MoPaCo and window conditions. Subjective evaluations showed similar assessment results for conversations and communication smoothness, suggesting MoPaCo usage results in smooth conversation and transmission of indications. In other words, this suggests indication actions were smoothly referenced by presenting through-window images, considering size and positional relationships in the media space as if the two spaces were actually joined by a glass window. From the experiment participants’ activities, we consider two reasons contributed to this.

Since MoPaCo presented spaces while preserving the geometrical consistency of width and positional relationships, gestures made at objects within the space and gaze targets were correctly transmitted. Under the 2D conditions, listeners would mistakenly look in the opposite direction of that being indicated, and indicators were often forced to name the object or otherwise provide concrete supplementary information. Figure 8 shows an example of this; the indicator (participant A) pointed to a clapperboard behind and to the right of the listener (participant B) using a gesture and gaze while saying, “Over there’s a thing making a clapping noise, what’s that called, a clapperboard?” (7m7s947 from conversation start). Participant B gazes at participant A and quickly identifies the target object, but the indicated direction is not transmitted directly, and participant B gazes in the opposite direction. Participant B says to participant A, “Which one?” (7m11s410). Participant A explains the location of the clapperboard in detail, saying “That thing that goes clap in TV and movies” (7m13s245). After that, participant B finally directs his gaze at the clapperboard, and says “Oh” while making a gesture (7m13s840). In this manner, since the direction of an indicator’s gaze and gestures cannot be accurately transmitted under the 2D conditions, finding an indication object often requires confirmation. Conversely, under the MoPaCo and window conditions, this sort of confirmation is not required. In other words, we consider using MoPaCo to reproduce the size and positional relationships of a space enables gaze and gesture directions to be accurately transmitted, allowing indication work to progress smoothly as if through an actual glass window.

We consider that through the window metaphor, since

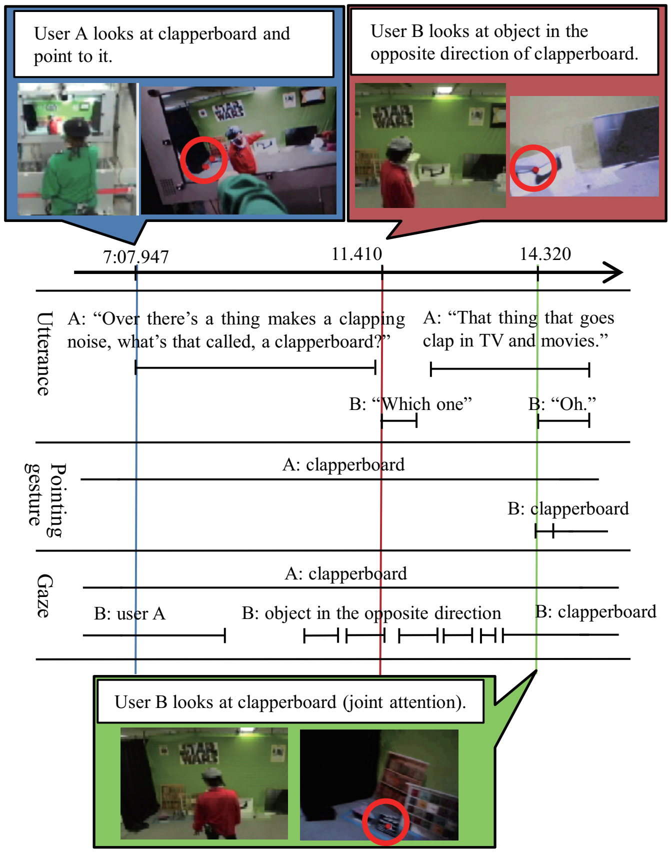


Figure 8: Example scene of instructor's action and recognition in 2D situation.

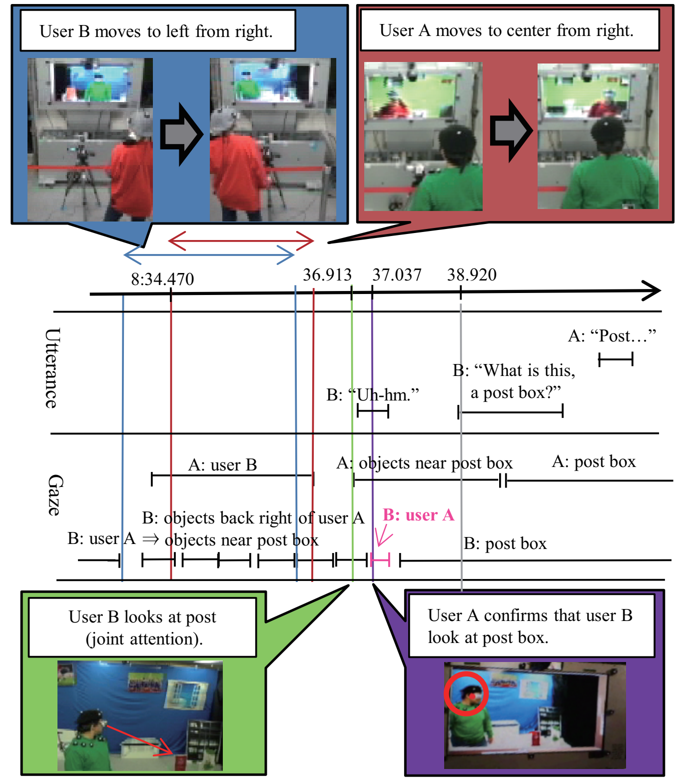


Figure 9: Example scene of instructor's action and recognition in MoPaCo situation.

gazing at objects was accompanied by a physical movement made by the user, changes in the user's position and direction were clear, and the partner could easily predict the object of the user's interest. Under the 2D conditions, since the entire room was displayed, users would move only their gaze without changing the position or direction of their head when gazing at an object. This made it difficult to grasp the direction of their gaze in the video, and participants were not often seen matching the gaze direction of their partner, moving their bodies in the same direction, or sharing mutual gazes. In contrast, under the MoPaCo and window conditions, when a participant gazed at an object, this was accompanied by a change in physical position or direction in most cases. The partner would then often change his or her position or direction to match the gaze. An example of this behavior under the MoPaCo conditions is shown in Figure 9. Participant B observes items in participant A's space from right to left. User A gazes at user B, and when he notices this movement, he moves from right to left to match user B's movement so he can always be seen from user B's position (8m34s470). When user B stops moving, user A also stops moving, turns his body toward the direction in which user B is looking, and shares a mutual gaze (8m36s913). In this case, we consider that user B is predicting the next instruction or explanation. Generally, user B directs his or her gaze at user A, and confirms user A is looking in the same direction in which he himself or she herself is looking (confirming he or her is sharing a mutual gaze) (8m37s037). He or she then indicates the post box they are both observing and says "What is this, a post box?" (8m38s920). At this point, user B's use of the demonstrative pronoun "this" indicates user B shares an interest target with user A and is predicting the next instruction or explanation. This sort of predictive behavior was seen under both MoPaCo and window conditions. Under both conditions, it was possible for users to firmly express targets (directions) of interest by observing objects using the glass window as a metaphor, which can be considered as a cause for this behavior. In other words, we consider users would show greater movements of the direction of their head and body by peeking through the glass window at something, allowing the partner to predict their target of interest (in the example shown in Figure 9, user A moves in response to user B's movements, and performs a mutual gaze). Since MoPaCo presented a window, when observing a space with a degree of size such as the one used in the experiment in this study, not all items could be observed at once, and participants were forced to move. However, it was shown nonverbal communication transmission was smoother than it is when simply displaying a 2D image in which the entire room could be seen. Thus, if the objective is to allow collaborative work using indicative actions to be performed smoothly, it is important to allow the natural transmission of nonverbal information such as gazes and gestures to be performed even if the entire room cannot be seen at all times. From this viewpoint, the MoPaCo window interface can be considered effective. Improvements to enjoyment and affinity seen in subjective evaluations are thought to be secondary to the improvement in smooth nonverbal communication. Moreover, increases in memory show MoPaCo gives more impressive images and possibly has the effect of making users engage more actively in conversation.

 ISBN: 978-1-61208-325-4

VII. Conclusion

This paper described evaluations of our proposed MoPaCo window interface system, which allows the size and positional relationships of two remote spaces to be reproduced using one stationary camera. The results obtained in implementing the system and performing evaluation experiments on it show it allows gazes and pointing gestures to be transmitted in a similar way to transmitting them through an actual glass window. We also performed experiments to determine whether indicative actions, which are important in performing remote indicative work, could be smoothly referenced with the system. Experiment results suggest MoPaCo users can accurately identify target objects as if under face-to-face conditions through an actual glass window. Results of experiments on conversation quality show the system facilitates smooth conversation and communication and strengthens memories of the conversations, suggesting users actively engage in conversation and the system makes a strong impression on them.

References

[1] A. H. Anderson, E. G. Bard, C. Sotillo, G. Doherty-Sneddon, and A. Newlands, “The effects of face-to-face communication on the intelligibility of speech,” Perception and Psychophysics, 59, 1997, pp. 580–592.

[2] M. Argyle and J. Graham, “The Central Europe Experiment - looking at persons and looking at things,” Journal of Environmental Psychology and Nonverbal Behavior, 1, 1977, pp.6–16.

[3] N. Suzuki et al., “Nonverbal behaviors in cooperative work: a case study of successful and unsuccessful team,” Proceedings of the 29th Annual Conference of the Cognitive Science Society, 2007, pp. 195–196.

[4] C. Heath and P. Luff, “Disembodied Conduct: Interactional Asymmetries in Video Mediated Communication,” Proceedings of ACM Conference on Human Factors in Computing Systems, 1991, 99–103.

[5] R. Ishii, S. Ozawa, H. kawamura, and A. Kojima, “MoPaCo: high telepresence video communication system using motion parallax with monocular camera,” IEEE International Workshop on Human-Computer Interaction: Real-time vision aspects of natural user interfaces (ICCV Workshops), 2011, pp. 463–464.

[6] R. Ishii, S. Ozawa, T. Mukouchi, and N. Matsuura, “MoPaCo: pseudo 3D video communication system,” Proceedings of the 1st international conference on Human interface and the management of information: interacting with information - Volume Part II, 2011, pp. 131–140.

[7] S. Whittaker, “Theories and methods in mediated communication,” The Handbook of Discourse Processes, New Jersey: Erlbaum, 2003, pp.253–293.

[8] Y. I. Nakano, G. Reinstein, T. Stocky, and J. Cassell, “Towards a model of face-to-face grounding,” Proceedings of the 41st Annual Meeting on Association for Computational Linguistics (ALC’03), 2003, pp. 553–561.

[9] P. Auer, “Projection in interaction and projection in grammar,” Text, 25, 1, 2002, pp. 7–36.

[10] C. Goodwin, “Conversational organization: Interaction between speakers and hears,” Academic Press, New York, 1981.

[11] C. Gale and A. F. Monk, “Where am I looking? The accuracy of video-mediated gaze awareness,” Perception and Psychophysics, 62, 2000, pp. 586–595.

[12] T. V. Crater, “The picturephone system: service standards,” Bell System Technical Journal, 50, 1971, pp. 235–269.

[13] E. D. Mynatt,, J. Rowan, S. Craighill, and A. Jacobs, “Digital family portraits: supporting peace of mind for extended family members,” Proceedings of Conference on Human-Factors in Computing Systems, 2001, pp. 333–340.

[14] Jiejie Zhu, Ruigang Yang, and Xueqing Xiang, “Eye contact in video conference via fusion of time-of-flight depth sensor and stereo,” Journal of 3D Research, 2,3, 2011.

[15] S. M. Anstis, J. W. Mayhew, and T. Morley, “The perception of where a face or television portrait is looking,” American Jounal of Psychology, 82, 4, 1969, pp. 474–489.

[16] A. Sellen, B. Buxton, and J. Arnott, “Using spatial cues to improve videoconferencing,” Video proceedings of Conference on Human-Factors in Computing Systems (CHI), 1992, pp. 651–652.

[17] S. Tanaka, K. Okada, S. Kurihara, and Y. Matsushita, “Desktopconferencing System Using Multiple Still-Pictures: Desktop-MAJIC,” Proceedings of the ninth conference on European Conference on Computer Supported Cooperative Work (CSCW), 1996, pp.5–6.

[18] R. Vertegaal and Y. Ding, “Explaining effects of eye gaze on mediated group conversations: amount or synchronization?”, Proceedings of the 2002 ACM conference on Computer supported cooperative work, 2002, pp. 41–48.

[19] R. Vertegaal, I. Weevers, C. Sohn, and C. Cheung, “GAZE-2: conveying eye contact in group video conferencing using eye-controlled camera direction,” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2003, pp. 521–528.

[20] K. Okada, F. Maeda, Y. Ichikawa, and Y. Matsushita, “Multiparty video-conferencing at virtual social distance: MAJIC design,” Proceedings of the ninth conference on European Conference on Computer Supported Cooperative Work (CSCW), 1994, pp. 383–393.

[21] D. T. Nguyen and J. Canny, “MultiView: improving trust in group video conferencing through spatial faithfulness,” Proceedings of Conference on Human-Factors in Computing Systems (CHI), 2007, pp. 1465–1474.

[22] H. Ishii and M. Kobayashi, “ClearBoard: a seamless medium for shared drawing and conversation with eye contact,” Proceedings of Conference on Human-Factors in Computing Systems (CHI), 1992, pp. 525–532.

[23] H. Kuzuoka, “Spatial workspace collaboration: a sharedView video support system for remote collaboration capability,” Proceedings of ACM Conference on Human Factors in Computing Systems, 1992, pp. 533–540.

[24] Tobii glass, [retrieved: http://www.tobiiglasses.com/scientificresearch/, January, 2014]

[25] M. Kipp, “Anvil - a generic annotation tool for multimodal dialogue,” Proceedings of the 7th European Conference on Speech Communication and Technology, 2005, pp. 1367–1370.

[26] Z. Zhang, “A flexible new technique for camera calibration,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), 2000, pp. 1330-1334.

Smartphone-Based 3D Navigation Technique for Use in a Museum Exhibit

L.-P. Bergé[1], G. Perelman[1], M. Raynal[1], C. Sanza[1], M. Serrano[1], M. Houry-Panchetti[1], R. Cabanac[2], E. Dubois[1]

[1] IRIT, Université Paul Sabatier
University of Toulouse
Toulouse, France
{Louis-Pierre.Berge, Emmanuel.Dubois}@irit.fr

[2] Télescope Bernard Lyot, Observatoire Midi-Pyrénées
University of Toulouse
Tarbes, France
remi.cabanac@obs-mip.fr

***Abstract*—3D Virtual Environment (3DVE) comes up as a good solution for transmitting knowledge in a museum exhibit. However, interaction techniques involved in such settings are mostly based on traditional devices such as keyboard and mouse. Recently, the popular use of smartphone as a personal handled computer lets us envision the use of mobile device as an interaction support with these 3DVE. In this paper, we focus on the navigation task inside a 3DVE and we propose to use the smartphone as a tangible object. Physical actions on the smartphone trigger translations and rotations in the 3DVE. In order to prove the interest in the use of the smartphone, we compare our solution with available solutions: keyboard-mouse and 3D mouse. User experiments confirmed our hypothesis and particularly emphasizes that visitors find our solution more attractive and stimulating.**

Keywords-interaction with smarpthone, 3D navigation, museum exhibit, experiment.

I. INTRODUCTION

With the evolution of technology, computing capabilities and rendering techniques, the use of 3D Virtual Environments (3DVE) is becoming a standard. They are no longer restricted to industrial use and they are now available to the mass-market in various situations: for leisure in video games, to explore a city in Google Earth or in public displays [1], to design a kitchen on a store website or to observe rare or fragile objects in a museum. However, in these mass market contexts, the visitor's attention must be focused on the content of the message and not distracted by any difficulties caused by the use of a complex interaction technique. This is especially true in a museum where the maximization of the knowledge transfer is the primary goal of an interactive 3D experience. Common devices, such as keyboard and mouse [2] or joystick [3] are therefore widely used in museums. To increase the immersion of the user, solutions combining multiple screens or cave-like devices [4] also exist. However, these solutions are cumbersome, expensive and not widespread.

Alternatively the use of smartphone, as a personal handheld computer, is commonly and largely accepted. Smartphones provide a rich set of features and sensors that can be useful to interact, especially with 3DVE and with remote, shared and large displays. Smartphones also create the opportunity for the simultaneous presence of a private space of interaction and a private space of viewing coupled with a public viewing on another screen. Furthermore, many researches have already been performed with smartphone to study their use for interacting with a computer. They explore multiple aspects such as technological capabilities [5], tactile interaction techniques [6], near or around interaction techniques [7]. Given the potential in terms of interaction support and the availability of smartphones in anyone's pocket, we explore in this paper benefits and limitations of the use of a smartphone for interacting with a 3DVE displayed in a museum context.

This paper is focusing on one task: the navigation inside a 3DVE. This is the most predominant task a user will have to perform in order to discover and understand the virtual space. Concretely our technique translates motions of the smartphone into motions of the point of view in the 3DVE. We thus propose 1) to consider a smartphone as a tangible object, in order to smoothly integrate it in a museum environment and because it has been proven to be easier to apprehend by newcomers [8], 2) to display feedback and/or personalized information on the smartphone display, 3) to deport the display of the 3DVE on a large screen, in order to provide a display space visible by multiple users as required in museum contexts. As a result, our technique combines the use of a popular and personal portable device, the physical space surrounding the device and the user gestures and input for navigating inside a 3DVE.

We compared our designed solution to the use of more common and available technologies: the keyboard-mouse device and a 3D mouse device. We proposed a controlled evaluation focused on the interaction techniques: out of a specific museum context, the user will not be distracted by pedagogical content. We measure usability and attractiveness in conjunction with performance considerations. The results confirm the interest of considering the use of personal mobile devices for navigating inside a 3DVE: results are particularly significant in terms of user attractivity.

In the following sections, we first detail our interaction technique. Next, we present the settings of the user experiment. We finally discuss the results, the place of our solution in the related work and we conclude with perspectives for improving the technique.

II. OUR INTERACTION TECHNIQUE

As described above, our interaction technique is based on the manipulation and use of a smartphone, a familiar and personal object for most of the users. Three major characteristics define our interaction technique: tangible manipulation of the smartphone, personalized data displayed on the smartphone and 3DVE displayed on a remote screen.

 ISBN: 978-1-61208-325-4

We restricted the degrees of freedom of the navigation task in order to be close to human behavior and existing solutions in video game with standard device: two degrees of freedom (DOF) are used for translations (front/back and left/right) and two for rotations (up/down and left/right). We did not include the y-axis translation and the z-axis rotation since they are not commonly used for the navigation task. To identify how to map the tangible use of the smartphone to these DOF, we first performed a guessability study.

A. Guessability study

14 participants have been involved and they were all handling their own smartphone in the right hand. To facilitate the understanding, the guessability study dealt with only one translation and one rotation. A picture of a 3DVE was presented to participants on a vertical support. It included a door on the left of the 3D scene and participants were asked to perform any actions they wished with their smartphone in order to be able to look through the door. A second picture was then displayed: now facing the door, participants were instructed to pass through the door.

In this second question, 11 participants performed hand translations to translate the point of view. Interestingly none suggested using the tactile modality. Results are more contrasted with the first question, requiring a rotation: 5 participants used a heading rotation of the handled smartphone; 1 only used the roll technique; 3 proposed to touch the target with their smartphone; 5 participants placed the smartphone vertically (either in landscape or portrait orientation) and then rotated the smartphone according to the vertical axis (roll) thus preventing the view on the smartphone screen.

B. Design solution

From the guessability study, we retain that physical translations of the smartphone seem to be the most direct way to perform translations of the point of view in the 3DVE. It has been implemented in our technique as follow. Bringing the smartphone to the left / right / front or back of its initial position triggers a corresponding shifting movement of the point of view in the 3DVE (Figure 1-a). The position of the point of view is thus controlled through a rate control approach; the rate applied is always the same and constant. In addition, feedback is provided on the smartphone. A large circle represents the initial position of the smartphone and the physical area in which no action will be triggered: the neutral zone. A small circle represents the current position of the smartphone and arrows express the action triggered in the 3DVE. As long as the small circle is inside the large circle, the navigation in the 3DVE is not activated. In addition it is possible to combine front / back translation with right / left translation. The feedback provided during each of four possible motions is illustrated in Figure 1-b. Finally, the smartphone vibrates every time that the navigation action is changed.

Regarding rotations, we retain from the guessability study the most usable solution: rotations of the hand-wrist handling the smartphone are mapped to orientations of the point of view in the 3DVE. In our implementation, horizontal wrist rotations to the left/right of the arm are mapped to left/right rotations of the viewpoint (heading axis, rY) and wrist rotations above/below the arm are mapped to up/down rotations of the viewpoint (pitch axis, rX) (Figure 1-c). A position control approach has been adopted here that establishes a direct coupling of the wrist angle with the point of view orientation. A constant gain has been set for the wrist rotations: the limited range of 10° to left and right [9] can be used to cover the range of the rotation angle inside the 3DVE (180°). This solution does not support a U-turn: this was not required in the experiment but could be solved by transforming the position control into a rate control when the wrist reaches a certain angle. In addition, two "spirit levels" feedback are displayed on the smartphone to provide an estimation of the current orientations of the smartphone (Figure 1-d) with respect to the initial orientations used as a reference.

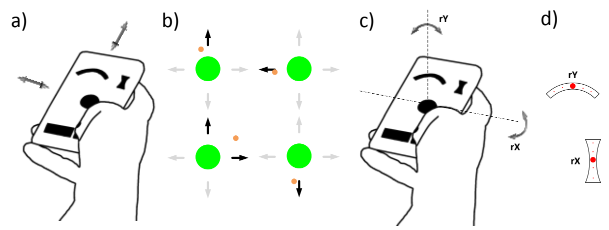


Figure 1: Our smartphone based interaction. (a) Physical action for translation. (b) Feedback of the translation: front, left, front and right, back translation. (c) Physical action for rotation. (d) Feedback of the rotation.

To avoid unintended motions of the virtual camera in the 3DVE, translations and rotations of the smartphone are applied to the 3DVE only when the user is pressing a button "navigate" displayed on the smartphone.

The smartphone also displays a "calibrate" button. This allows the user to recalibrate the smartphone at will, i.e., to reset the center of the neutral zone to the current position of the smartphone and the reference orientations.

III. EXPERIMENT

We conducted an experiment to compare our smartphone-based interaction technique with two other techniques using devices available in museums: a keyboard-mouse combination and a 3D mouse. In the museum context, the temporal performances are not predominant. In fact our goal was to assess and compare the usability and attractiveness of these three techniques. Our protocol does not include museum information in order to keep the participant focused on the interaction task.

A. Task

The task consisted in navigating inside a 3D tunnel composed by linear segments ending in a door (Figure 2-b). The task is similar to the one presented in [10] and sufficiently generic to correctly evaluate the interaction techniques. Participants had to go through the segments and go across the doors but could not get out of the tunnel. Black arrows on the wall allow finding easily the direction of the next door. The segments between the doors formed the tunnel, whose orientation was randomly generated in order to include all 25 possible directions to the next door. The center of each door is placed 0, 20 or 40 pixels to the left or right and to the top

 ISBN: 978-1-61208-325-4

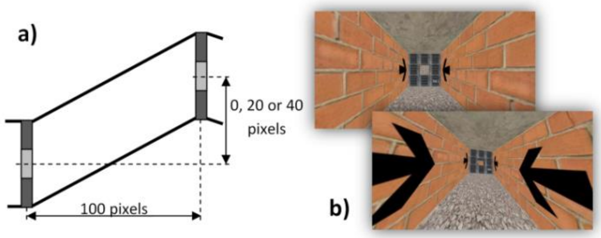


Figure 2. (a) Representation of one segment of the 3D tunnel. (b) Screenshots of the 3D environment of the experiment.

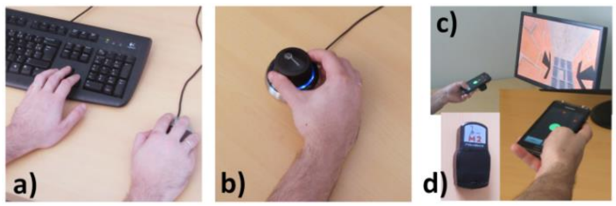


Figure 3. (a) The keyboard-mouse, (b) The 3D mouse and (c) the smartphone configuration. (d) The Polhemus sensor.

or bottom of the center of the previous door (Figure 2-a). The movement of the user is not subject to gravity. When the user looks up and starts a front translation movement, the resulting motion is a translation in the direction of point targeted.

B. *Interaction techniques*

We compared three techniques: keyboard-mouse, 3D mouse and our technique based on a smartphone. In keyboard-mouse, movements of the mouse control the 2 DOF point of view of the virtual camera (orientation). The four directional arrow of the keyboard control the 2 DOF of the translation of the virtual camera. In 3D mouse, the participant applies lateral forces onto the device to control translation (right/left, front/back), and rotational forces to control orientations of the virtual camera. The use of our technique, the smartphone, has been described in section II. For the three techniques, it appears that left/right translations are particularly useful when collision with doors occurs.

For each technique we determined the speed gain of the translation and rotation tasks through a pre-experiment with six subjects. We asked participants to navigate inside our 3D virtual environment with each technique and to freely adjust the gains to feel comfortable performing the task. We stopped the experiment and recorded the settings when the participant successfully went through 5 consecutive doors. Finally, for each technique, we averaged the values of gain between participants. We noticed that the gain of the translation of keyboard-mouse was higher than the 3D mouse or smartphone. This is probably due to the habit of subjects to manipulate this technique.

C. *Apparatus*

The experiment was done in full-screen mode on a 24″ monitor with a resolution of 1920 by 1080 pixels. We developed the environment with a 3D open source engine, Irrlicht, in C++. For the keyboard and mouse device, we used a conventional optical mouse and a standard keyboard with 180 keys (Figure 3-a). For the 3D mouse we used the Space-Navigator [11] (Figure 3-b), a commercial device with 6 DOF. For the smartphone, we implemented the technique on a Samsung Galaxy S2 running Android 4.1.2 (Figure 3-c). To avoid an overload of the smartphone computing capabilities with the processing of the internal sensors (accelerometers, gyroscope) we used an external 6D tracker: the Polhemus Patriot Wireless (Figure 3-d). We attached a sensor on the rear face of the smartphone. Via a driver written in C++, the marker returns the position and the orientation of the smart-phone. We filtered the data noise with the 1€ filter [12].

D. *Participants and procedure*

We recruited a group of 24 subjects (6 female), of 29.3 (SD=9) years old on average. All subjects were familiar with the keyboard and mouse, 17 of them had a smartphone and only 1 has used the 3D mouse.

Every participant performed the 3 techniques (smartphone, keyboard-mouse and 3D mouse). They started with the keyboard-mouse technique in order to be used as a reference. The order of smartphone and 3D mouse techniques was counterbalanced to limit the effect of learning, fatigue and concentration. For each technique, the subject navigated inside 6 different itineraries. We counterbalanced the itineraries associated with each technique across participants so that each technique was used repeatedly with each group of users.

Participants were sited during the experiment and were instructed to optimize the path, i.e., the distance travelled. They could train themselves on each technique through one itinerary. When the user passed through a door, a positive beep was played. When the user collided with an edge of the tunnel, a negative beep was played.

After having completed the six trials for one technique, the subject filled the SUS [13] and AttrakDiff [14] questionnaires and indicated three positive and negative aspects of the technique. The procedure is repeated for the two remaining technique. The experiment ended with a short interview to collect oral feedback. The overall duration of the experiment was about 1 hour and 30 minutes per participant.

E. *Collected data*

In addition to the SUS and AttrakDiff questionnaires filled after each technique to measure usability and attractiveness, we also asked for a ranking of the three interaction techniques in terms of preferences. From a quantitative point of view we measured the traveled distance and the number of collisions.

IV. RESULTS

We present in the following section quantitative and qualitative results obtained.

A. *Quantitative results*

First a Kruskal-Wallis test confirmed that none of the 18 randomly chosen itineraries had an influence on the collected results. On average we observed that the travelled distance is the smallest with the keyboard-mouse (2766*px*, SD = 79), followed by the 3D mouse (2881*px*, SD = 125) and the smartphone (2996*px*, SD = 225). According to a Wilcoxon test these differences are significant. The same conclusions

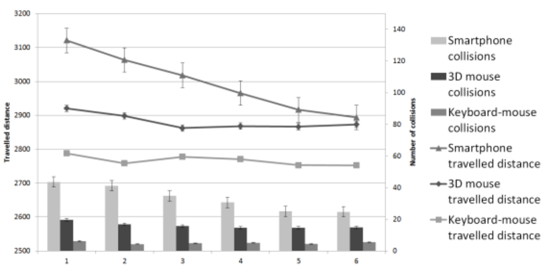


Figure 4: Evolution of the travelled distance and the amount of collisions according to 6 trials of the subjects

Figure 5: Portfolio generated on the AttrakDiff website

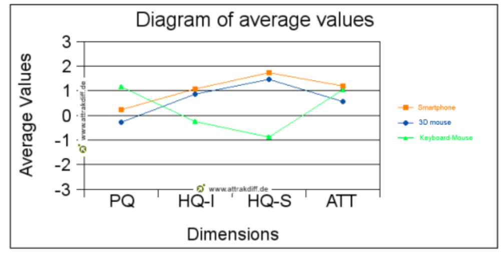


Figure 6: Average values for the four dimensions of the AttrakDiff questionnaire

can be drawn with regard to the amount of collisions (keyboard-mouse: 5.08, SD = 5.68; 3D mouse: 16.11, SD = 15.86; smartphone: 33.35, SD = 24.64).

Given the high dispersion of the distance and collision measures, we refined this analysis in distinguishing the results obtained for each of the six trials performed by the 24 participants (Figure 4). This refined analysis reveals a significant learning effect with the smartphone technique: between the first and sixth trial, the distance is 7.3% shorter (Wilcoxon test, $p = 6 \times 10^{-6}$) and collision are reduced of 43.3% (Wilcoxon test, $p = 2 \times 10^{-4}$). A significant learning effect is also observed with the 3D mouse, but only in terms of distance and with a smaller improvement (1.6% shorter, Wilcoxon test, $p = 0.049$).

The learning effect with the smartphone is so important that, at the last trial, the travelled distance for the smartphone (2893*px*) and the 3D mouse (2873*px*) is comparable (no significant difference, Wilcoxon test, $p = 0.49$).

B. *Qualitative results*

Three aspects have been considered in the qualitative evaluation: usability, attractiveness and user's preference.

Usability evaluation: the SUS questionnaire [13] gives an average score of 82.60 (SD=12.90) for the keyboard-mouse, 54.79 (SD=22.47) for the smartphone based interaction and 53.54 (SD=27.97) for the 3D mouse. A Wilcoxon test shows that the SUS difference is statistically significant between the keyboard-mouse and each of the two other techniques (3D mouse, smartphone). However, the SUS difference is not statically significant between the 3D mouse and the smartphone. Research conducted on the interpretation of the SUS score [15] permits to classify the usability of the keyboard-mouse as "excellent". According to this same interpretation scale, the usability of the smartphone and the 3D mouse is identified as "ok".

We also note a wide dispersion of the SUS score. We thus performed a more detailed analysis of the SUS score. First, according to [15] a system with a "good" usability must obtain a score above 70. In our experiment, 33% of participants scored the 3D mouse above 70 while 37% of participants scored the smartphone above 70. Second, 3D mouse and smartphone were two techniques unfamiliar to the participants. Results of the SUS questionnaire shows that when the smartphone is used after the 3D mouse, the average score for the smartphone is 65.62 whereas in the other order the average score is 43.96. The perceived usability of the two unfamiliar techniques is therefore lower than the perceived usability of the keyboard-mouse. However, once the participants have manipulated these two unfamiliar techniques, the perceived usability of the smartphone increases drastically.

Attractiveness: Data collected using AttrakDiff [16] give an idea of the attractiveness of the technique and how it is experienced. Attrakdiff supports the evaluation of a system according to four distinct dimensions: the pragmatic quality (PQ: product usability, indicates if the users could achieve their goals using it); the hedonic quality – stimulation (HQ-S: determine to which extent the product can support the need in terms of novel, interesting and stimulating functions, contents and interaction); the hedonic quality – identity (HQ-I: indicates to what extend the product allows the user to identity with it); the attractiveness (ATT: global values of the product based on the quality perception).

Figure 5 shows a portfolio of average value of the PQ and the HQ (HQ-S+HQ-I) for the three interaction techniques assessed in our user experiment.

The keyboard-mouse was rated as "fairly practice-oriented", i.e., one of the first levels in the category "task-oriented". According to the website report [16], the average value of PQ (above 1) indicates that there is *definite* room of improvement in terms of usability. The average value of HQ obtained (approx. -1) expresses that there is *clearly* room for improvement in terms of user's stimulation. The 3D mouse was rated as "fairly self-oriented", i.e., one of the first levels in the category "self- oriented". The average value of PQ (approx. 0) expresses that there is room for improvement in terms of usability. The average value of HQ obtained (approx. 1) expresses that room for improvement also *exists* in terms of user's stimulation. The smartphone was rated as "self-oriented". The average value of PQ (approx. 0) expresses that there is room for improvement in terms of usability. The average value of HQ obtained (above 1) expresses that the user identifies with the product and is motivated and stimulated by it.

Figure 6 summarizes the average values for the four AttrakDiff dimensions of the three interaction techniques. With regards to the four dimensions the smartphone is rated higher than the 3D mouse and the differences are statistically significant (T-test, p<0.05). For the PQ value the keyboard-mouse is better than the smartphone (statistically significant, p<0.05). For HQ-I and HQ-S values the smartphone is better than the keyboard-mouse (statistically significant, p<0.05).

In terms of ATT, smartphone is again rated higher than keyboard-mouse but the difference is however not statistically significant ($p>0.05$). Compared to the keyboard-mouse, the smartphone is considered as novel, innovative, inventive, stylish and creative. Improvements in term of simplicity, straightforwardness or predictability could increase the average value of PQ and probably increase even more the ATT value of the smartphone.

User preference: at the end of the experiment a short semi-guided interview was performed. Participants were first asked to rank the three techniques from 1 (best) to 3 (worst). Results are in line with the SUS scores: the keyboard-mouse technique is largely preferred and the 3D mouse is by far the least appreciated technique: only 2 participants out of 24 ranked it as the best, and 14 ranked it as the worst. The smartphone based-interaction is ranked uniformly in the three places (7, 9, 8).

Finally, three positive points and three negative points were asked for each technique. The most frequently mentioned positive points are "quick", "easy" and "accurate" for the keyboard-mouse technique, "intuitive", "novel" and "usable with on hand" for the 3D mouse and "immersive", "funny", and "accessible to everybody" for the smartphone. Participants are thus appreciating the conditions of use created by the smartphone while they particularly pinpoint the effectiveness of the mouse and provide general comments about the 3D mouse.

The most frequently mentioned negative aspects is a practical aspect for the keyboard-mouse ("requires the use of both hands"). They are related to the effectiveness of use of the 3D mouse ("difficulty to combine translation and rotation at the same time", "lack of precision" and "high need for concentration") and for the smartphone it focuses on one specific feature ("difficulty to translate to the left or right") and the overall context of use ("the apparent time of learning" and "the fatigue caused in the arm").

Technical issues for the 3D mouse and effectiveness of the keyboard-mouse are thus highlighted while benefits and limits related to the interactive experience are mentioned for the smartphone. This clear shift of interest between the three techniques reveals that the disappointing performances of the smartphone highlighted in the previous section are not totally overruling the interest of the participants for the smartphone-based technique. It is therefore a very interesting proof of interest for further exploring the use of smartphone in 3DVE.

V. DISCUSSION

Among the existing attempts for exploring the navigation of 3DVE with a smartphone, two different settings exist. A first set of solutions, as opposed to our setting, propose to display the 3DVE directly on the smartphone. Different techniques are explored to change the point of view inside the 3D scene: tactile screen like Navidget [17], integrated sensor [18], smartphone motions in the space around a reference, as Chameleon technique [19] and T(ether) [20] and manipulation of physical objects around the smartphone [21]. The second set of solutions avoids issues related to occlusion of the 3DVE with fingers by displaying the 3DVE on a distant screen. These solutions involved integrated sensor [22] to detect user's motions, tactile screen [23] or a combination of both [24]. Although our technique is clearly in line with this second set of solutions, our use of the smartphone presents three major originalities. Firstly, the smartphone is not limited to a remote controller: it is also used to provide the user with feedback or personalized information. Secondly, using tactile interaction to support the navigation would occlude part of the screen and prevent its use for visualizing data, selecting objects or clicking on additional features. Instead, physical gesture are applied to the smartphone to control rotations like in [25] or [24] but also to control translations of the point of view in the 3DVE. Thirdly, the choice of the gestures to apply has been guided by the results of a guessability study that highlight the most probable gesture users would perform with a smartphone. We used this approach rather than a pre-experiment or results of existing experiment [24] because when getting familiar with the manipulation of smartphone, universal gestures will be adopted, and not necessarily those known as the most efficient. User's prime intuition of use looked more important to us.

Beyond the interaction technique designed, the contribution includes a set of evaluation results. The user experiment revealed a significant learning effect with the smartphone. This is a very encouraging result because no learning effect was observed with keyboard-mouse and 3D mouse although the participants were unfamiliar with 3D mouse and smartphone: the use of smartphone thus significantly improves over the time. Results also revealed that the use of our smartphone based technique to navigate inside a 3DVE is more attractive and stimulating than more usual technique such as the keyboard-mouse and the 3D mouse. In terms of usability, user's preferences (interaction technique ranking) and quantitatively (travelled distance and amount of collisions) our smartphone technique appears to be weaker than the keyboard-mouse technique but similar to the 3D mouse. This tradeoff between attractivity and usability /performance emphasizes that compared to two manufactured devices, our technique is better accepted but weaker in performance and usability. This is particularly encouraging because technological improvements of our technique, such as mixing the use of integrated sensor with image processing to compute more robust and accurate smartphone position and orientation, will also increase the user's performance. In addition the use of smartphone is already widely spread and we believe that their use as an interaction support with remote application will develop as well and become a usual interaction form. Altogether this user experiment establishes that the use of smartphone to interact with 3DVE is very promising and need to be further explored.

Finally, to validate the interest of the approach in an operational context, the presented smartphone technique is currently running in an animated 3DVE representing a large telescope. The interactive installation (Figure 7) is used in a museum to explain the different parts of the telescope and how the telescope is operated. Visitors virtually navigate in the dome of the telescope to observe the different elements and perform complementary actions such as selecting a star or controlling the telescope. A large and varied audience,

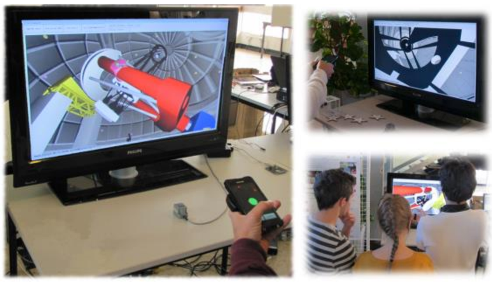


Figure 7: Interactive installation with our smartphone based interaction

ranging from scholar to retired persons, are using this interactive installation. An in-situ evaluation of the use of the technique in comparison to other will soon be performed.

VI. CONCLUSION

In this paper, we explored the feasibility of using a smartphone to navigate inside a 3DVE displayed in the context of a museum exhibit. Smartphones present the advantage to provide a private space for viewing and to constitute a personal device for navigating or controlling a 3D cultural or pedagogical content. Generalizing its use throughout a museum is also completely imaginable. With a QR code the visitor can easily download the mobile app in front of the exhibit and interact with the 3DVE. The originality of our technique relies on the fact that the smartphone is used as a tangible object. Physical actions on the smartphone trigger translation and rotation in the 3DVE. Very promising results have been highlighted in a user experiment comparing our solution to a keyboard-mouse technique and a 3D mouse, the most common devices found in museum nowadays. We measured that after a short learning time, the smartphone technique leads to performance results that are comparable to the 3D mouse. Through technical optimization we are also confident that it might become comparable to the keyboard-mouse technique. But more notably, we clearly established that visitors find such a solution more attractive and stimulating.

In this study, we have therefore established that the use of a smartphone as a tangible object for navigating inside a 3DVE is a good alternative to the keyboard-mouse and 3D mouse. A part from the technical improvements already mentioned we plan to investigate the causes and length of the learning effect. To complete our study, it could be interesting to measure the impact of the use of a smartphone on the museum visit and on the quality of the educational transfer. Finally, using a hand free interaction with the smartphone for navigating inside a 3DVE opens up perspectives for controlling additional features in the 3DVE with the smartphone that we will integrate.

REFERENCE

[1] D. S. Tan, D. Gergle, P. Scupelli, and R. Pausch, “Physically large displays improve performance on spatial tasks,” *ACM Trans. Comput. Interact.*, vol. 13, Mar. 2006, pp. 71–99.

[2] L. Pecchioli, M. Carrozzino, F. Mohamed, M. Bergamasco, and T. H. Kolbe, “ISEE: Information access through the navigation of a 3D interactive environment,” *J. Cult. Herit.*, vol. 12, no. 3, Jul. 2011, pp. 287–294.

[3] T. Wischgoll and J. Meyer, “An explorational exhibit of a pig’s heart,” in *ACM SIGGRAPH 2005*, 2005, p. 138.

[4] C. Christou, C. Angus, C. Loscos, A. Dettori, and M. Roussou, “A versatile large-scale multimodal VR system for cultural heritage visualization,” in *VRST’06*, 2006, pp. 133–140.

[5] H. Graf and K. Jung, “The smartphone as a 3D input device,” in *2012 IEEE Second International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, 2012, pp. 254–257.

[6] H.-N. Liang, J. Trenchard, M. Semegen, and P. Irani, “An exploration of interaction styles in mobile devices for navigating 3d environments,” in *APCHI’12*, 2012, pp. 309–313.

[7] D. Avrahami, J. O. Wobbrock, and S. Izadi, “Portico: tangible interaction on and around a tablet,” in *UIST'11*, 2011, pp. 347–356.

[8] O. Shaer and E. Hornecker, “Tangible User Interfaces: Past, Present, and Future Directions,” *Found. Trends® Human–Computer Interact.*, vol. 3, no. 1–2, Jan. 2009, pp. 1–137.

[9] T. Tsandilas, E. Dubois, and M. Raynal, “Modeless Pointing with Low-Precision Wrist Movements,” in *INTERACT 2013*, 2013, vol. 8119, pp. 494–511.

[10] G. Casiez, and C. Chaillou, “Effects of DOF Separation on Elastic Devices for the Navigation in 3D Virtual Environments with Force Feedback,” in *IEEE World Haptics 2005*, 2005, pp. 483-486.

[11] “3Dconnexion.” [Online]. http://www.3dconnexion.fr/products/spacenavigator. [Accessed: 27-Jan-2014].

[12] G. Casiez, N. Roussel, and D. Vogel, “1 € filter: a simple speed-based low-pass filter for noisy input in interactive systems,” in *CHI’12*, 2012, pp. 2527–2530.

[13] J. Brooke, “SUS: A quick and dirty usability scale,” *Usability Eval. Ind.*, 1996, pp. 189–194.

[14] M. Hassenzahl, “The Interplay of Beauty, Goodness, and Usability in Interactive Products,” *Human-Computer Interact.*, vol. 19, no. 4, Dec. 2004, pp. 319–349.

[15] A. Bangor, P. T. Kortum, and J. T. Miller, “An Empirical Evaluation of the System Usability Scale,” *Int. J. Hum. Comput. Interact.*, vol. 24, no. 6, Jul. 2008, pp. 574–594.

[16] “AttrakDiff.” [Online]. http://attrakdiff.de/index-en.html. [Accessed: 27-Jan-2014].

[17] M. Hachet, F. Decle, S. Knodel, and P. Guitton, “Navidget for 3D interaction: Camera positioning and further uses,” *Int. J. Hum. Comput. Stud.*, vol. 67, no. 3, 2009, pp. 225–236.

[18] W. Hürst and M. Helder, “Mobile 3D graphics and virtual reality interaction,” in *ACE’11*, 2011, pp. 28–36.

[19] B. Buxton and G. W. Fitzmaurice, “HMDs, Caves & Chameleon: A Human-Centric Analysis of Interaction in Virtual Space,” *SIGGRAPH*, vol. 32, no. 4, 1998, pp. 64–68.

[20] “Tangible Media Group.” [Online]. http://tangible.media.mit.edu/project/tether/. [Accessed: 27-Jan-2014].

[21] M. Hachet, J. Pouderoux, and P. Guitton, “3D Elastic Control for Mobile Devices,” *IEEE Comput. Graph. Appl.*, vol. 28, no. 4, Jul. 2008., pp. 58–62.

[22] S. Boring, M. Jurmu, and A. Butz, “Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays,” in *OZCHI’09*, 2009, pp. 161–168.

[23] D. Gracanin, K. Matkovic, and F. Quek, “iPhone/iPod Touch as Input Devices for Navigation in Immersive Virtual Environments,” in *IEEE Virtual Reality Conference*, 2009, pp. 261–262.

[24] A. Benzina, A. Dey, M. Toennis, and G. Klinker, “Empirical evaluation of mapping functions for navigation in virtual reality using phones with integrated sensors,” in *APCHI’12*, 2012, pp. 149–158.

[25] F. Daiber, L. Li, and A. Krüger, “Designing gestures for mobile 3D gaming,” in *MUM’12*, 2012, pp. 3–8.

A 3D Interface to Explore and Manipulate multi-scale Virtual Scenes using the Leap Motion Controller

Bruno Fanini
CNR ITABC
Rome, Italy
bruno.fanini@gmail.com

Abstract — **Gesture-based interaction models can be efficient and simpler to understand if designed to correspond to common user interactions with the physical world. This paper presents a 3D interface and its implementation to quickly perform navigation and manipulation tasks in multi-scale and multi-resolution 3D scenes using a low-cost consumer sensor: the Leap Motion controller. The developed system has the goal of exploring the potential of accurate hands and fingers tracking alongside mid-air 3D gestures, to investigate specific design advantages and issues they present in such complex environments.**

Keywords — Gesture-based interaction; 3D manipulation; real-time; multi-scale

I. Introduction and Related Work

Interaction and navigation within complex 3D virtual contexts, in conjunction with a well-designed graphical user interface (GUI), are crucial to final user experience. Mouse-based GUIs have proven their robustness and flexibility, but a major shift towards "natural" user interfaces (NUIs) during the latest years is strong: this is not only related to research, but to applications and tools targeting broader consumer audiences as well.

A 3D interface is "*a User Interface that involves 3D interaction*" and 3D interaction can be defined as a layer that allows the user to perform different tasks *directly* in a 3D spatial context [9]. A widely used approach for 3D interfaces involves a physical 3D space used as input for the application: user provides such input by making gestures in this space. Typically, the software application is equipped with a gesture recognition system through the use of a sensor (e.g., Kinect [7] and others) allowing to map such physical movements into a set of predefined functionalities and actions. Unlike classic devices such as keyboard or mouse, these sensors allow to *spatially track* specific joints of a human body (e.g., arms, shoulders, hands, fingers, etc.) within a 3D context, providing for instance 3D positions <x,y,z>, motion data and orientations of different features. On the application side, the developer has to provide a software layer to process incoming data and recognize specific patterns (gestures) over an observed time slice, transforming these spatial inputs into direct actions inside the virtual 3D environment.

These devices and their progress in sensor accuracy, speed and efficiency [2][19] during the last years are laying solid foundations to deploy astonishing interaction models. Furthermore, low-cost systems and devices to detect hand [8] and body gestures [7] are nowadays becoming widely available to consumer market. Obviously, there are limitations of such tracking controllers that have to be considered for the design of efficient interaction models, such as device accuracy, noise issues and lighting conditions.

Some of the goals of 3D interfaces are to increase user engagement (for instance within serious gaming applications), application usability (natural mapping from 3D physical space to a 3D virtual space) and even reducing a few common bottlenecks related to 3D-oriented tasks [12]. Since 3D interaction is a quite recent topic, the maturity of 3D interface design principles lags behind those for standard GUIs. Given the wide range and diversity of input devices and interaction models, there isn't actually an established standard for 3D User Interfaces. While general Human Computer Interaction (HCI) principles such as Nielsen's heuristics [1] still apply, they are not sufficient for designing a usable 3D user interface.

This paper will describe a 3D interface design and its software implementation applied to a recent consumer device: the Leap Motion controller. The developed system is designed to explore and manipulate 3D objects in real-time within multi-scale and multi-resolution 3D virtual environments, using OpenSceneGraph framework as 3D visualization front-end. The following two sections will briefly introduce the device (II) and the OpenSceneGraph framework (III) while Section IV will describe the 3D interface and a test case where the system was applied.

II. The Leap Motion controller

The Leap Motion controller [8] is a small and inexpensive motion sensor available to consumer market since July 2013, composed of 2 small cameras and 3 infrared LEDs (Light-Emitting Diode) able to track hands, fingers and a few tools in mid-air inside a specific field of view, with sub-millimeter accuracy [2]. This 8 cm long USB (Universal Serial Bus) peripheral device (Figure 1) is designed to be placed on a physical desktop, facing upward. It operates in an intimate proximity, with a very high tracking frame-rate (when lighting conditions are optimal) inside a field of view of the shape of an inverted pyramid, centered on the controller.

 ISBN: 978-1-61208-325-4

Figure 1. The Leap Motion controller.

Several tests on the device report an effective range of 25 mm to 600 mm above the controller, where hands, fingers and tools are detected in mid-air with 3D positions, gestures, and other motion data. The growing community of developers around the Leap Motion has access to the latest Software Development Kit (SDK), Application Programming Interface (API) documentation and forums. The provided API allows to listen and report real-time data regarding hands (palm position, normal, direction, etc.) fingers (tip position, direction, etc.) and tools, alongside a few built-in recognized gestures such as circle, swipe, tap gestures, and a few others.

III. The OpenSceneGraph framework

OpenSceneGraph (OSG) [14] is an open source 3D *rendering middleware* and one of the world's leading scene graph API [3] used by application developers in fields of visual simulation, computer games, virtual reality, scientific visualization and modeling. The OpenSceneGraph framework is widely used within real-time 3D scientific visualization contexts due to its performance, portability and scalability, providing a huge set of functionalities that won't be discussed in detail in this paper. It is based on scene-graph structures, thus allowing the definition of spatial and logical relationships among different 3D models (nodes) in a virtual scene, specifically efficient on large and multi-resolution datasets. It allows to develop real-time 3D applications providing:

- Object-oriented functionalities
- Transformation nodes
- Loading of common 3D formats (Alias Wavefront OBJ [16], Autodesk 3D Studio Max [17], COLLADA [18], etc.)
- Management of large 3D environments, using spatial segmentation of the virtual world
- Remote node loading (via URL)
- Efficient management of level of detail (LoD)
- Instancing
- Paging

Transformation node in OSG is particularly useful to manage an object disposition, since it encapsulates a matrix transformation (position, rotation, scale) that is being applied to the entire sub-graph. Multi-resolution datasets are also fully supported and use appropriate representation (Level of Detail) depending on current camera view, while instancing techniques are able to reduce memory footprint through node sharing. Paging mechanisms in OpenSceneGraph allow scene portions (or "*pages*") to be loaded and unloaded at run-time from the main scene-graph, reducing system workload and GPU load, depending on current point of view and frustum.

Figure 2. Some examples of large-scale and multi-resolution 3D scenes in real-time using Front-Ends based on OpenSceneGraph framework.

Several test cases and projects (Figure 2) demonstrated the framework efficiency and performance specifically in handling and visualizing large, complex and multi-resolution datasets such as terrains, cities, etc. [4][5]. These are a few reasons that led to the framework choice as Front-End of 3D visualization in the current implementation of the proposed interface.

 ISBN: 978-1-61208-325-4

IV. 3D Interface Design and Implementation

This section will describe the 3D interface with its components, functionalities and overall design for navigation and manipulation in a multi-scale, multi-resolution 3D virtual environment using the Leap Motion controller. The proposed 3D interface is based on a two-handed input design [11][15]: in fact, one of the goals is to provide an efficient and fast interaction model for 3D-oriented tasks [12].

The whole concept takes advantage of the device accuracy, although several tests during design and development of first prototypes exhibited some data noise in special conditions, such as when hands approach the sides of the controller field-of-view. These issues were mostly solved at software level by applying special smoothing filters to received data, in this case to the features specifically involved in the design, such as palms and fingers. For completeness, a few definitions are provided:

1. *Virtual World (W)* encapsulates the scene-graph of the world (the whole scene or "global space"), potentially composed by complex/multi-scale 3D datasets.
2. *Manipulables* represent a collection of nodes (3D objects) subset of the virtual world W, having the property of being "editable" and able to be transformed over the time.
3. *View Configuration (v)* is composed by position (eye: $<x,y,z>$) and orientation (quaternion: $<x,y,z,w>$) representing a camera view or point of view into the current virtual world.
4. *Transition:* $v_t \rightarrow v_{t+1}$ where v is a View Configuration that changes through user interaction.
5. *Interaction Space (IS)* represents the 3D manipulation domain. It's located into the global space (W) through a transformation T_{is} (position, rotation and scale) and maps the physical space above the Leap controller.

It is important to mention that virtual world *W* and *manipulables* collection can rely both on local and remote locations (or mixed) allowing very interesting scenarios for the interface applied to any OpenSceneGraph-based application.

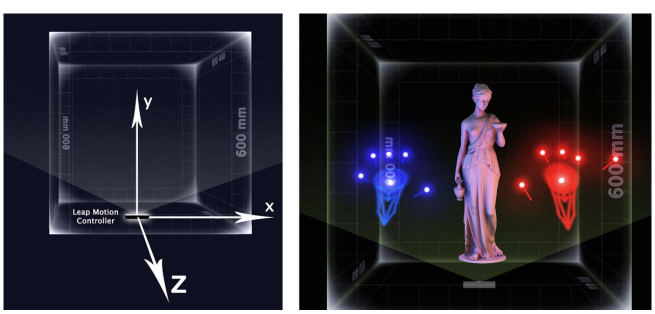


Figure 3. Interaction Space (IS) coordinate system and 3D visualization of both hands, fingers and device field-of-view (bottom).

The implemented system consists of a real-time visualization *Front-End*, a gesture listener and a 3D scene manager. Recognized features by the device, such as hands and fingers, are visualized and directly mapped inside the IS using a right-handed Cartesian coordinate system (Figure 3) with origin centered on the middle IR LED of the device. The rightmost hand in the developed system is represented in red while the leftmost hand in blue. Both hands are luminescent to provide customized real-time lighting of 3D models. The IS cube (600 x 600 x 600 mm) represents a reference of the physical space above the controller inside the virtual space.

The system provides the fundamental functionality to map local IS coordinates $<x,y,z>$ (Leap Space) to World coordinates $<x_w,y_w,z_w>$ (World Space) and viceversa (1), for instance mapping a fingertip into world (global scene) coordinates.

$$<x_w,y_w,z_w> = T_{is}(<x,y,z>) \quad (1)$$

The next sub-section will describe the navigation model, how it influences T_{is} and when.

A. Navigation

A fulfilling exploration of a virtual world is a complex topic and requires special attention. Using standard peripherals, there are several well-established and familiar navigation models, for instance *pan-rotate-zoom* using the mouse, just to name one. When dealing with large virtual environments, there are even additional issues that need to be addressed. Content rarefaction in particular, is a typically disorienting aspect, although some solutions based on hotspots affordance have been proposed to reduce this phenomenon [13]. In such context, a 3D interface can provide a more efficient navigation model, allowing to combine different actions at the same time using a single gesture (e.g., pan action + zoom action, etc.) although sensor accuracy clearly plays a crucial role in this scenario. In general, each exploration session *E* can be represented as:

$$v = E(t). \quad (2)$$

where v is a varying *View Configuration* over time, depending on user input. Let us suppose the user starts from a large-scale context, wanting to focus on a detail of a small-scale 3D object to perform some task: how much time is required before the user reaches a satisfying view configuration v_f ? One of the goals of the proposed 3D interface is to minimize the t in (2) required for v_f.

The developed navigation model consists of two distinct phases: *stop* and *drag*. When the user is in stop phase, the system simply visualizes both hands and fingers inside the Interaction Space, listening for gestures. The user has a direct mapping of his own hands inside the manipulation domain since the IS is fixed in global space W (T_{is} is not changed) and aligned with the current View Configuration (fixed). When the drag gesture is recognized, user is able to fully manipulate current View Configuration by "*dragging*" the space above the Leap Motion controller using both hands simultaneously, leading to a fluid bi-manual camera control model [6] in 3D. Previous work in literature [15] indeed suggests that specific tasks related to two-handed input can be performed effectively with a symmetric assignment of roles to both hands: in this case, camera motion and target are controlled by the position of both palms in 3D space. The system allows to combine into a single gesture a 5-*DoF* camera manipulation and *3D scale*. Denoting P_{left} and P_{right} as positions <x,y,z> of both palms in IS and C as the center between the two palms:

- *Left/Right, Up/Down* and *Forward/Back* are controlled by corresponding position of C relatively to IS.
- *Yaw* is controlled by palms difference along the IS z-axis.
- *Pitch* is controlled by average palms normal.
- *3D Scale* is controlled by palms distance and Target spot. This feature, similar to the *pinch* gesture on 2D multi-touch devices, provides the multi-scale exploration.

Scale manipulation feedback is provided by the radius of a ring overlay element that shows the relative scaling factor being applied and the target spot C', according to the center of the ring C. During the whole phase, IS is not changed: this design provides the user with a visual reference of the old IS state, useful in a multi-scale context since it enhances scale perception during camera transitions. For instance, increasing palms distance when a drag gesture is recognized, will shrink the IS and consequently scale the virtual hands in comparison to the world scene W.

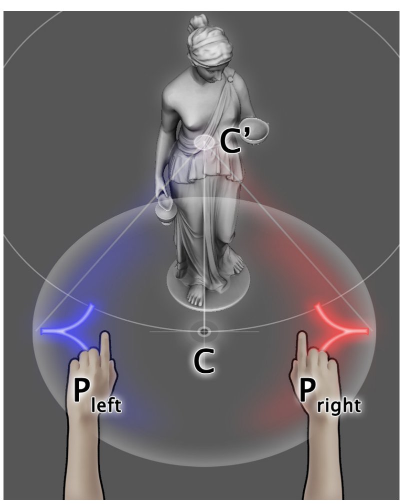


Figure 4. Drag gesture and target spot C' used in navigation model.

When dragging gesture ends (release), a 3D target spot C' is directly picked through C using an intersection ray on 3D geometry, as illustrated in Figure 4: the target spot is used to provide a new *View Configuration* and thus a new suitable IS. The system performs a validation of maximum and minimum length values of the segment $\overline{C\text{-}C'}$ to limit maximum and minimum scale, respectively. At this point, the user is once again in a stop phase and can iterate the whole process. Figure 5 illustrates a fast sequence of drag and stop to reach a partially occluded spot in a multi-resolution 3D model: starting from a stop phase (1) with no hands above the controller and a sample 3D model inside the IS, a *drag* gesture is initiated (2).

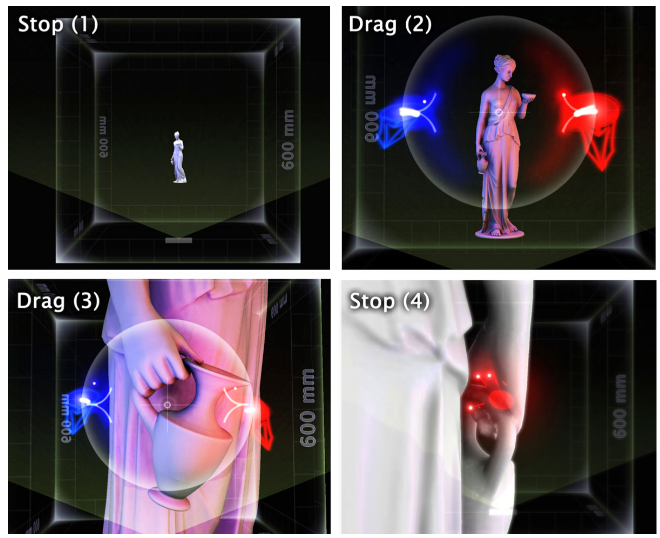


Figure 5. Stop and drag navigation model.

Notice how in (3) the center point of the ring is used to aim at the desired 3D target spot and in (4) the rightmost hand is partially intersecting domain geometries during a stop phase inside a new IS, with different scale. The cumulative nature of this approach aims to provide precise control by taking advantage of the Leap Motion device sub-millimeter accuracy [2], quickly alternating *stop* and *drag* phases to reach complex View Configurations in short times, maximizing user control during a transition $v_t \rightarrow v_{t+1}$. The scheme shown in Figure 6 is a graph of stop and drag sequences recorded during a test session on a sample 3D scene and their relation with Interaction Space scale magnitude over time. Stop sequences have variable durations since user is observing scene details, operating on IS domain (further details in the next Section) or resting (hands exit the device field-of-view). The phase alternation has also the goal of reducing fatigue for arms and hands [9][10]. The graph focuses on scale magnitude: notice how T_{is} is computed only at the beginning of a stop phase and at the end of a drag phase. In the recorded test-case, user session started from a scene overview (a) with IS mapped on a scale magnitude of 10 mt, moving in to focus on a detail (b) at a scale magnitude of 1 cm, then scaling up over one meter magnitude (c) and then focusing on a different 3D detail on 1 mm scale (d). The system also allows to keep track of Interaction Spaces over the session: this is also useful to roll back to the previous IS with a simple gesture.

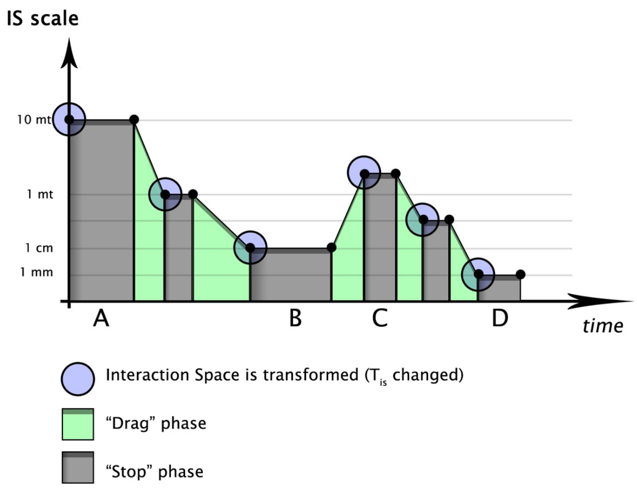


Figure 6. Recorded graph of Interaction Space magnitude (scale) over time.

The user is also able to record a specific View Configuration v_r for later recall, so that current pair <v_r , *scale*> can be stored on demand. The existence of *stop* phase in the proposed interface is further explained in the following sub-section.

B. Manipulation

When the user is not adjusting the view configuration through a *drag* phase, the system is in *stop* phase: hands are free to move inside the Interaction Space and what is currently contained. The design is specifically conceived to leave room for object manipulation tasks and other operations within current domain. During the stop phase, the user can activate an *edit* mode: this mode allows to apply several actions to items belonging to *manipulables* collection. Once the edit mode is activated by using a specific gesture and a visual feedback, user is able to operate on current domain and 3D objects inside the IS applying different transformations. Red hand (*dominant*) is used to grab a 3D object and move it using the direct IS mapping from physical space above the device and the virtual world. A selector is used when user is hovering a manipulable item to provide visual feedback. Object manipulation is thus performed by grabbing the virtual item and by moving it within the current Interaction Space (and its current scale). Grabbing gesture is recognized by the system on highlighted item when hand contraction is over a specified threshold and the item is picked up. When detected inside the IS, the blue hand (*non-dominant*) is used to apply additional transformations (rotation, scale, etc.) to the current selection. Like the proposed navigation model, in this case, bi-manual editing allows to apply multiple transformations at the same time (e.g., rotating and translating a picked object). The user can also take advantage of the multi-scale concept of the interface by getting closer to a spot and by operating on 3D objects using fine grained transformations. For instance, moving an entire house over a multi-resolution terrain and then zooming in to place a spoon on a table, since transformation accuracy is tied to IS scale.

C. Basic arrangement experiment

After the initial prototype, a test case was created using a basic scene consisting of a ground and a sample collection of *manipulables* (Figure 7): table, chair, armchair and a lamp. In order to test and verify basic usability of the system, a few internal lab sessions were carried out: users were asked to create plausible items arrangements through the developed system, starting from a randomized disposition (a). Initial observation sessions related to manipulation tasks provided sufficient data to modify the visual feedback through a white selector (b), that is shown only when the user hovers the item using the dominant hand (red). While the grabbed item is moved, the user can apply yaw rotation using non-dominant hand (c), finally reaching a plausible disposition of the scene (d). A second test case was performed to test navigation and manipulation of remote 3D datasets: the user was able to operate on grass, flowers and trees, streamed over the net (Figure 8).

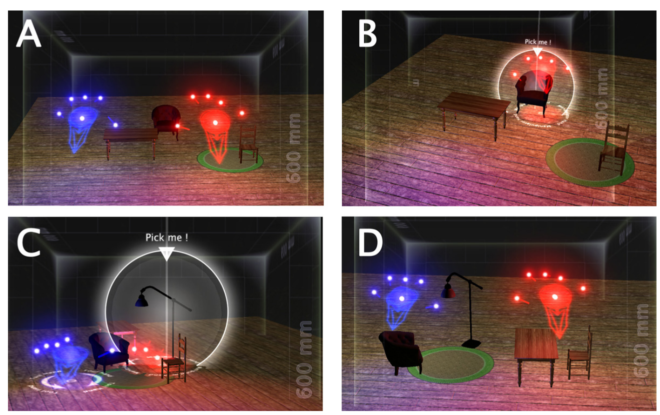


Figure 7. Test case for manipulation of 3D objects inside the Interaction Space.

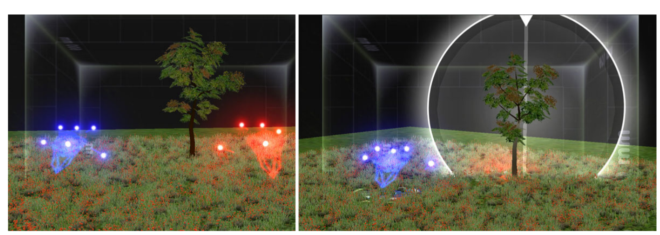


Figure 8. Manipulation and navigation of remote 3D datasets

These internal experiments and observations allowed to test and increase basic usability of the system by:

- Adding visual feedback on manipulable hovering (3D overlay selector).
- Adding constraints to navigation by limiting minimum-maximum IS scale inside a specific range and by enclosing allowable target spots (drag phase) inside global *manipulables* bounding-box.
- Adding explanatory welcome panels.

The same tasks were performed on the test scene using a mouse-based interface, providing participants with basic translation and rotation features.

V. Conclusion and Future Work

A 3D interface and its implementation were presented to explore and manipulate multi-scale and multi-resolution virtual environments. Although few experimental tests and observations were carried out on a limited number of users to evaluate interface usability, the following results can be highlighted:

- A navigation and manipulation model operating on multi-scale virtual environments.
- Efficient and interactive visualization on large, complex and even remote 3D datasets provided through the *OpenSceneGraph* framework integration [3][4].
- Use of a small and inexpensive desktop motion sensor (Leap Motion controller).
- A *Stop* and *Drag* design that maximizes the controller accuracy through relative *view configuration* transitions (Drag phase) and reduces arms fatigue during user sessions (Stop phase).
- Quick, multi-scale bi-manual manipulation (*Edit mode*) within the current IS through direct mapping.

The developed system that implements the 3D interface is based on the Leap Motion API and the OpenSceneGraph framework, allowing to be easily deployed to consumer market since the availability of the Leap Motion controller to the public. In order to obtain more accurate usability results, an adequate number of participants will be tested and more than one scene along with complex manipulation tasks will be provided. The integration with the OpenSceneGraph framework also provides portability across different Operating Systems (Windows, Linux, MacOS) and also inherits a large set of features provided by the framework, such as paging capabilities introduced in Section III for management of large scenes and manipulation of remote 3D scenes or sub-graphs (server-based approach).

The concept and design of Interaction Space allows great scalability and further extension. For instance, attached toolboxes to Interaction Space have been tested to provide the user with a set of selectable functionalities (duplicate object, remove object, etc.). Further advancements will be oriented to extend the manipulation functionalities and to introduce more constraints, such as axis snapping. The effectiveness of these additions will be supported by user testing, observations and experiments.

References

[1] J. Nielsen and R. Molich, “Heuristic evaluation of user interfaces”, Proceedings of CHI 90, pp. 249-256. New York, NY: ACM, 1990.

[2] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler, "Analysis of the accuracy and robustness of the Leap Motion controller", Sensors 2013, 13, pp. 6380-6393, ISSN 1424-8220, 2013.

[3] W. Rui and Q. Xuelei, "OpenSceneGraph 3.0", Packt Publishing 2010 (eBook).

[4] B. Fanini, L. Calori, D. Ferdani, and S. Pescarin, "Interactive 3D landscapes online", International Archive of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVIII-5/W16, 2011.

[5] X. Li, D. Kong, Y. Zhang, and Y. Sun, "Mass Models Dynamic Loading Method Research and Implementation based on OpenSceneGraph", Journal of Information & Computational Science 8: 2 (2011), pp. 362-369, 2011.

[6] R. Balakrishnan and G. Kurtenbach, "Exploring Bimanual Camera Control and Object Manipulation in 3D Graphics Interfaces", Proceedings of the ACM CHI 99 Human Factors in Computing Systems Conference, pp. 56-63, 1999.

[7] Microsoft Kinect (Windows): http://www.microsoft.com/en-us/kinectforwindows. Last accessed: 24th November 2013

[8] Leap Motion controller: https://www.leapmotion.com. Last accessed: 24th November 2013

[9] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev, "3D User Interfaces – Theory and Practice". Addison Wesley Longman Publishing Co., Inc., Redwood City, CA (USA), 2005.

[10] M. R. Mine, "Virtual environment interaction techniques", Tech. report, Chapel Hill, NC (USA), 1995

[11] W. Buxton and B. Myers, "A study in two-handed input", Proceedings of CHI 1986, New York, NY (USA) ACM, pp. 321-326, 1986.

[12] M. W. Gribnau and J. M. Hennessey, "Comparing single- and two-handed 3D input for a 3D object assembly task", Proceedings of CHI 1998, New York, NY (USA) ACM, pp. 233-234, 1998.

[13] B. Fanini, "Development and Evolution of Natural Interfaces in Virtual Heritage Applications", MIMOS Proceedings, 2012

[14] OpenSceneGraph: http://www.openscenegraph.org. Last accessed: 24th November 2013

[15] R. Balakrishnan and K. Hinckley, "Symmetric bimanual interaction", Proceedings of the SIGCHI conference on Human Factors in Computing Systems, ACM New York, NY (USA), ISBN:1-58113-216-6, pp. 33-40, 2000

[16] Alias/Wavefront OBJ file format: http://www.martinreddy.net/gfx/3d/OBJ.spec

[17] 3D Studio Max file format: http://www.fileformat.info/format/3ds/egff.htm

[18] COLLADA 3D model format: https://collada.org/

[19] K. Khoshelham, "Accuracy Analysis of Kinect Depth Data", International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXVIII-5/W12, 2011

 ISBN: 978-1-61208-325-4

Touchscreen User Motion Anticipation for Usability Improvement

Tetsuyou Watanabe and Sawako Furuya
School of Mechanical Engineering
Kanazawa University
Kanazawa, Japan
emails: {te-watanabe@ieee.org, Sawakofuruya14@gmail.com}

***Abstract*— This paper proposes a method for improving touchscreen usability by anticipating user motions. Reaction speed and simple structure are important for good touchscreen usability. With this in mind, we present a system that can estimate a desired position for user motion by anticipating a motion several time steps ahead, using only sensors attached to the touchscreen. User motion that changes suddenly may not have the Markov property. We present here a novel methodology based on an Auxiliary Particle Filter (APF) with parameter estimation to deal with this issue. User motion is anticipated by regarding the motion in terms of parameters. We demonstrate the validity of our approach through experiments.**

Keywords-Auxiliary particle filter; anticipation; user intention; touchscreen; table device; user interface

I. INTRODUCTION

Many machine products are currently available, most of which require human operation. A machine that requires human operation does not always run as intended. Touchscreen manipulation is a typical example. A user can control items displayed on a touchscreen using fingers in an intuitive manner. However, a system sometimes cannot react to rapid finger motion, resulting in the display of an unintended image. Anticipating user motion to produce an output corresponding to user intention can be expected to result in improved work efficiency and reduced human error.

The important issues in user interface design are product downsizing (a hardware issue) and reaction speed (a software issue). A large user interface system can have reduced mobility, difficult set up, and therefore limited use. A late system response might not always produce the output that a user wants. Developing a small system for a user interface with rapid reaction can be expected to enhance convenience and usage in various fields, such as medicine. With this in mind, this paper targets the development of a convenient user interface system for touchscreens. By predicting intended user motion, namely, the desired position for the motion, we speed up the reaction speed of a system and enhance its usability. We use only regular sensors originally attached to a touchscreen to construct a compact system.

There are some research on predicting user intention. Chen et al. [1] proposed a method for predicting user intention/action in web applications using a user intention model based on extracted linguistic features. Armentano et al. [2] developed a method in which interface agents can detect user intention based on a variable-order Hidden Markov Model (HMM). Carrasco et al. [3] presented a method for predicting user hand movement using information from a camera attached to the wrist. Surf features extracted from the image were used to predict hand motion using an HMM. However, their target was *sequential* user intention or action expressed using discrete state variables, and the presented methods are difficult to apply to a continuous process such as user touchscreen motion. However, the predicted motion is a manually discrete event. There are several methods using a Particle Filter (PF) [4] for tracking humans and objects. Zahidul et al. [5] developed a PF-based method for tracking objects in a video scene using color and shape information. Wang et al. [6] presented a methodology for dealing with occlusion by combining the information from multiple cameras. The purpose of such research is not the anticipation of user motion (estimation of user intention), but its detection. Anticipation is required in situations in which a user motion suddenly changes, and its tracking is temporarily lost. A method that can deal with sudden changes in user motion is proposed by Oka et al. [7]. They developed a system that estimates the three-dimensional posture of a user's head in real time using PF from multiple camera images. However, this method did not provide an adequate solution. Enlarging the search area for a state at the next time step results in sudden changes in user motion from a discrete state change to a non-discrete (continuous) state change. This method thus requires a large number of particles and cannot be expected to work on touchscreens in real time.

On the other hand, research dealing with touchscreens has targeted mainly the evaluation of systems to improve designs. Wobbrock et al. [8] analyzed gestures used with touchscreens. Lin [9] evaluated the usability of mobile maps running on touchscreens. There is some research on developing systems for improving touchscreen usability. Dohse et al. [10] presented a method that combines hand tracking and touch detection to detect the movement of each user when multiple users access one touch display. However, a camera was used as an additional sensor for touch detection.

With this in mind, this paper presents a methodology for estimating user intention (desired position for user motion) for touchscreens. The key features are as follows. (1) Only regular sensors originally attached to a touchscreen are used. (2) We improve the reaction speed of a system by estimating user intention continually, using an Auxiliary Particle Filter (APF) [11]. APF is a kind of PF. PF is a Bayes filter [12] that approximates posterior density using a number of particles. Each particle has a weight, and the particles are recursively updated/sampled according to their weights. When computing weights in APF, observation results are taken into account. Particles closer to the actual value than in PF can be generated in APF. APF has been used for such tasks as robot localization [13], anomaly detection in spacecraft [14], and human tracking [15]. APF was extended to a filter that can deal with parameter estimation by West [16] and Liu et al. [17]. User intention (user-desired position) is usually continuous, but sometimes discontinuous. For example, a user's finger moves from left to right (in this case the motion is continuous), but the direction of motion changes suddenly from right to upward (in such a moment the motion is discontinuous). Then, user intention may lack the Markov property. This indicates that we cannot handle user intention estimation similarly to state estimation. In this connection, this paper presents a novel methodology that regards user intention in terms of parameters, and applies APF with parameter estimation.

The remainder of this paper is organized as follows. In Section II, the target system is described using a basic strategy. APF is introduced in Section III. We present our APF-based methodology for estimating user intention in Section IV. The experimental results are shown in Section V, and we conclude in Section VI.

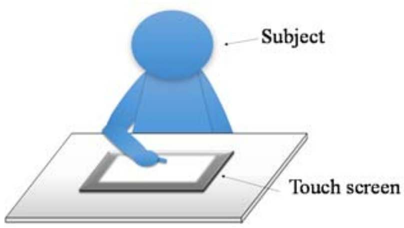


Figure 1. Schematic view of the target system

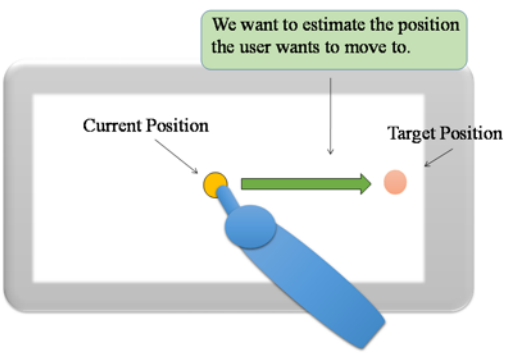


Figure 2. User operation on touch screen

II. TARGET SYSTEM AND BASIC STRATEGY

Fig. 1 shows a schematic view of the target system. We suppose that the user uses the touchscreen with one finger and moves object images on the touchscreen. Our purpose is to anticipate the position to which the user wants to move the target, namely, the user intention (Fig. 2).

The user finger position is measured by sensors attached to the touchscreen. This measurement is the observation value. We denote the observation value at time t by $\boldsymbol{y}_t \in R^2$. Let the user finger position be the state of the system: $\boldsymbol{x}_t \in R^2$. Ideally, $\boldsymbol{x}_t = \boldsymbol{y}_t$, but owing to (for example) noise and reaction delay, this does not always hold in practice. We regard user intention as the desired target position of user motion. As mentioned previously, the desired target position may lack the Markov property. We then regard the desired target position in terms of parameters: $\boldsymbol{\theta}_t \in R^2$. This is the key idea for estimating user intention.

In this case, state and observation equations can be represented by

$$p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{\theta}_t) \; , \tag{1}$$

$$p(\boldsymbol{y}_t|\boldsymbol{x}_t) \; , \tag{2}$$

with the aim of estimating user intention $\boldsymbol{\theta}_t$ while estimating the state $\boldsymbol{x}_t$:

$$p(\boldsymbol{x}_t, \boldsymbol{\theta}_t|\boldsymbol{y}_t) \; . \tag{3}$$

We use APF with parameter estimation [16][17] for this for this purpose. The original method estimates time-invariant parameters, but the target parameter $\boldsymbol{\theta}_t$ is not always time-invariant. Therefore, we extend the original APF with parameter estimation for estimating time-variant parameters.

III. AUXILIARY PARTICLE FILTER WITH PARAMETER ESTIMATION

We briefly introduce APF with parameter estimation as presented by West [16] and Liu et al. [17]. APF can be regarded as an extended version of PF based on the Sampling Importance Resampling (SIR) algorithm [18][19][20]. APF is explained after a brief introduction to PF.

A. Particle Filter based on SIR

PF is a Bayesian filter for estimating the state $\boldsymbol{x}_t$ at time t from observation $\boldsymbol{y}_t$. It corresponds to deriving the posterior distribution $(\boldsymbol{x}_t|\boldsymbol{y}_t)$. From Bayes' theorem, we have

$$p(\boldsymbol{x}_{t+1}|\boldsymbol{y}_{t+1}) \propto p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1}) p(\boldsymbol{x}_{t+1}) \; . \tag{4}$$

Here, the prior distribution can be expressed by

$$p(\boldsymbol{x}_{t+1}) = \int p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t) p(\boldsymbol{x}_t|\boldsymbol{y}_t) d\boldsymbol{x}_t \; . \tag{5}$$

PF is for approximating the posterior distribution $p(\boldsymbol{x}_t|\boldsymbol{y}_t)$ at each time step t by $i = 1, \ldots, N$ particles $\boldsymbol{x}_t^{(i)}$ with corresponding weights $\omega_t^{(i)}$. Then, (5) is rewritten as

$$p(\boldsymbol{x}_{t+1}) = \sum_{j=1}^{N} \omega_t^{(j)} p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t^{(j)}) \; , \tag{6}$$

after which (4) is rewritten as

$$p(\boldsymbol{x}_{t+1}|\boldsymbol{y}_{t+1}) \propto p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1}) \sum_{j=1}^{N} \omega_t^{(j)} p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t^{(j)}) \; . \tag{7}$$

We sample from $p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t^{(j)})$ for the j^{th} particle and evaluate the corresponding value of $\omega_t^{(j)} \, p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})$. Let this value be $\omega_{t+1}^{(j)}$ (at the next time step).

 ISBN: 978-1-61208-325-4

$$\omega_{t+1}^{(j)} \propto \omega_t^{(j)} p\left(\boldsymbol{y}_{t+1} \middle| \boldsymbol{x}_{t+1}^{(j)}\right) = \omega_t^{(j)} \frac{p(y_{t+1}|x_{t+1}^{(j)})p(x_{t+1}^{(j)}|x_t^{(j)})}{q(x_{t+1}^{(j)}|x_t^{(j)},y_t)} . \quad (8)$$

The second term is the general form from the viewpoint of importance sampling. In importance sampling, the state is expressed by the weighted sum of proposal distribution $q(\cdot)$. SIR-based PF can be regarded as the filter assuming that the proposal distribution $q(\cdot)$ is the prior distribution $p(\boldsymbol{x}_t)$.

In SIR-based PF, the sampled points depend on the current prior $\boldsymbol{x}_{t+1}$ and do not take $\boldsymbol{y}_{t+1}$. into account. Consequently, we need a large number of particles for accurate estimation. This is the issue in PF. APF is a filter that resolves this issue by taking the likelihood distribution $p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})$ into account.

B. Auxiliary Particle Filter

APF resolves the issue as follows.

$$p(\boldsymbol{x}_{t+1}|\boldsymbol{y}_{t+1}) \propto \sum_{j=1}^{N} \omega_t^{(j)} p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})\, p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t^{(j)}) . \quad (10)$$

The difference in the position of summation in (10) from its position in (7) indicates that in (10), $\boldsymbol{y}_{t+1}$, derived/estimated from the observation function, is counted when sampling particles in (10), but is not counted when sampling in (7). As a result, APF counts only meaningful particles, while PF has to count meaningless particles as well. This means that fewer particles are required in APF compared to PF. Improved accuracy is also expected in APF.

In practice, it is difficult to know $\boldsymbol{x}_{t+1}$ in $p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})$. directly. Then, for the j^{th} particle, $\boldsymbol{\mu}_{t+1}^{(j)}$ is considered instead of $\boldsymbol{x}_{t+1}$ in $(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1})$ **,** where $\boldsymbol{\mu}_{t+1}^{(j)}$ is a representative value, such as the mean of $p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t^{(j)})$. We evaluate the following weight for the j^{th} particle, instead of (8).

$$g_{t+1}^{(j)} \propto \omega_t^{(j)} p(\boldsymbol{y}_{t+1}|\boldsymbol{\mu}_{t+1}^{(j)}) , \quad (11)$$

where $\omega_t^{(j)}$ is updated at time $t + 1$ as follows.

$$\omega_{t+1}^{(j)} = \frac{p(y_{t+1}|x_{t+1}^{(j)})}{p(y_{t+1}|\mu_{t+1}^{(j)})} . \quad (12)$$

This indicates that the weight $\omega_t^{(j)}$ is updated such that for every particle, the posterior distribution given in (4) can be expected to be a maximum. Control of the diffusion of particles is also expected.

C. APF with Parameter Estimation

APF is extended by West et al. [16] to a filter that can estimate (time-invariant) parameters. Here the filter is introduced briefly.

Let $\boldsymbol{D}_t$ be all of the information that the system has at time t. We consider the case in which the parameters $\boldsymbol{\theta}_t$ are time-invariant: $\boldsymbol{\theta}_t = \boldsymbol{\theta}$. In this case, the posterior function is expressed by

$$p(\boldsymbol{x}_{t+1}, \boldsymbol{\theta}|\boldsymbol{D}_{t+1}) \propto p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1}, \boldsymbol{\theta}) p(\boldsymbol{x}_{t+1}, \boldsymbol{\theta}|\boldsymbol{D}_t)$$

$$\propto p(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1}, \boldsymbol{\theta}) p(\boldsymbol{x}_{t+1}|\boldsymbol{\theta}, \boldsymbol{D}_t) p(\boldsymbol{\theta}|\boldsymbol{D}_t) . \quad (13)$$

$\boldsymbol{\theta}$ cannot be treated in the same way as in state $\boldsymbol{x}_t$ estimation. Thus, we add small random disturbances and evolve artificially

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \boldsymbol{\zeta}_{t+1} , \quad (14)$$

$$\boldsymbol{\zeta}_{t+1} \sim N(0, \boldsymbol{W}_{t+1}) , \quad (15)$$

where $N(0, \boldsymbol{W}_{t+1})$ is the normal distribution with mean 0 and variance matrix $\boldsymbol{W}_{t+1}$, and $\boldsymbol{\zeta}_{t+1}$ provides small random disturbances. Then, $\boldsymbol{\theta}$ becomes $\boldsymbol{\theta}_t$ and can be treated in a manner similar to that used for state estimation.

Here, supposing that $\boldsymbol{\theta}_t^{(j)}$ is the parameter for the j^{th} particle with corresponding weight $\omega_t^{(j)}$, the following equation is considered for the parameter estimations.

$$p(\boldsymbol{\theta}_t|D_t) \approx \sum_{j=1}^{N} \omega_t^{(j)} N(\boldsymbol{\theta}_t|\boldsymbol{m}_t^{(j)}, h^2\boldsymbol{V}_t) , \quad (16)$$

where $N(\boldsymbol{\theta}_t|\boldsymbol{m}_t^{(j)}, h^2\boldsymbol{V}_t)$ is a multivariate normal density function with mean $\boldsymbol{m}_t^{(j)}$ and variance matrix $h^2\boldsymbol{V}_t$. h (> 0) is the smoothing parameter. $\boldsymbol{V}_t$ is the weighted variance matrix for $\boldsymbol{\theta}_t$.

$$\boldsymbol{V_t} = \sum_{j=1}^{N} \omega_t^{(j)} (\boldsymbol{\theta}_t^{(j)} - \overline{\boldsymbol{\theta}}_t)(\boldsymbol{\theta}_t^{(j)} - \overline{\boldsymbol{\theta}}_t)^T , \quad (17)$$

where

$$\overline{\boldsymbol{\theta}}_t = \sum_{j=1}^{N} \omega_t^{(j)} \boldsymbol{\theta}_t^{(j)} \quad (18)$$

$\boldsymbol{m}_t^{(j)}$ is the mean for $\boldsymbol{\theta}_t$ for the j^{th} particle and then, normally, $\boldsymbol{m}_t^{(j)} = \boldsymbol{\theta}_t^{(j)}$. However, in this case, the variance matrix for $p(\boldsymbol{\theta}_t|D_t)$ given in (17) becomes $(1 + h^2)\boldsymbol{V}_t$, and over-dispersion occurs. In order to resolve this issue, West [16] introduced the following equation for $\boldsymbol{m}_t^{(j)}$**.**

$$\boldsymbol{m}_t^{(j)} = a\boldsymbol{\theta}_t^{(j)} + (1 - a)\overline{\boldsymbol{\theta}}_t , \quad (19)$$

where $a = \sqrt{1 - h^2}$. In this case, the variance matrix for $p(\boldsymbol{\theta}_t|D_t)$ becomes $\boldsymbol{V}_t$ **,** and the issue is resolved.

APF with parameter estimation can be summarized as follows ([16] and [17]). Table 1 shows the nomenclature.

1. For $j = 1, \ldots, N$, we calculate the likely value associated with $\boldsymbol{x}_t^{(j)}$, $\boldsymbol{\theta}_t^{(j)}$, given by $\boldsymbol{\mu}_{t+1}^{(j)}$, $\boldsymbol{m}_t^{(j)}$

$$\boldsymbol{\mu}_{t+1}^{(j)} = E(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t^{(j)}, \boldsymbol{\theta}_t^{(j)}) , \quad (20)$$

$$\boldsymbol{m}_t^{(j)} = a\boldsymbol{\theta}_t^{(j)} + (1 - a)\overline{\boldsymbol{\theta}}_t , \quad (19)$$

where $E(X|A)$ is the conditional expectation of random variable X under phenomenon A. Note that a in (19) determines how close $\boldsymbol{\theta}_t^{(j)}$ should be to the mean $\overline{\boldsymbol{\theta}}_t$. In practice, a is set to a value extremely close to one (0.973-0.994).

2. Resample N particles according to the following weight.

$$g_{t+1}^{(j)} \propto \omega_t^{(j)} p(\boldsymbol{y}_{t+1} | \boldsymbol{\mu}_{t+1}^{(j)}, \boldsymbol{m}_t^{(j)}) \,. \tag{21}$$

Note that the index k is used for the resampled N particles.

3. Sample the parameters according to

$$\boldsymbol{\theta}_{t+1}^{(k)} \sim N(\cdot | \, \boldsymbol{m}_t^{(k)}, h^2 \boldsymbol{V}_t) \,. \tag{22}$$

4. Sample a value of the current state $\boldsymbol{x}_{t+1}^{(k)}$ from the state equation.

$$\boldsymbol{x}_{t+1}^{(k)} \sim p(\cdot \, | \boldsymbol{x}_t^{(k)}, \boldsymbol{\theta}_{t+1}^{(k)}) \tag{23}$$

5. Evaluate the following weight.

$$\omega_{t+1}^{(k)} \propto \frac{p(y_{t+1} | \, x_{t+1}^{(k)}, \boldsymbol{\theta}_{t+1}^{(k)})}{p(y_{t+1} | \, \mu_{t+1}^{(k)}, \boldsymbol{m}_t^{(k)})} \tag{24}$$

6. Repeat steps one through five.

TABLE I. NOMENCLATURE

N	The number of particles
$\boldsymbol{x}_t, \boldsymbol{y}_t, \boldsymbol{\theta}_t$	The state, the observation and the parameter vectors at t
$\boldsymbol{x}_t^{(j)}$	$\boldsymbol{x}_t$ at the j^{th} particle at t
$\boldsymbol{\theta}_t^{(j)}$	$\boldsymbol{\theta}_t$ at the j^{th} particle at t
$\omega_t^{(j)}$	Weight of j^{th} particle at t
$\overline{\boldsymbol{\theta}_t}$	$\sum_{j=1}^{N} \omega_t^{(j)} \boldsymbol{\theta}_t^{(j)}$
$\boldsymbol{V}_t$	$\sum_{j=1}^{N} \omega_t^{(j)} (\boldsymbol{\theta}_t^{(j)} - \overline{\boldsymbol{\theta}_t})(\boldsymbol{\theta}_t^{(j)} - \overline{\boldsymbol{\theta}_t})^T$
$\overline{\boldsymbol{\theta}_t}$	$\sum_{j=1}^{N} \omega_t^{(j)} \boldsymbol{\theta}_t^{(j)}$
$\boldsymbol{\mu}_{t+1}^{(j)}$	Likely value associated with the component $p(\boldsymbol{x}_{t+1} | \boldsymbol{x}_t^{(j)}, \boldsymbol{\theta}_t^{(j)})$
$\boldsymbol{m}_t^{(j)}$	Likely value associated with $\boldsymbol{\theta}_t^{(j)}$
$g_t^{(j)}$	Weight for resampling at the j^{th} particle at t

IV. User Intention Estimation Using Parameterization

Here we present a new methodology for estimating user intention. User motion can change suddenly, in which case the Markov property cannot be assumed and the usual state estimation algorithm cannot be used. The key idea for resolving this issue is to estimate the desired position for user motion in terms of parameters. We use APF with parameter estimation for user intention. The original method is intended for time-invariant parameters, while the target parameters are not always time invariant. However, we can use the same analogy, with some extensions of the algorithm, because the method treats the time-invariant parameters as time-variable parameters.

The desired position for user motion $\boldsymbol{\theta}_t^u$ is considered to be sometimes the same as or close to the value at the previous time step and sometimes definitely different and difficult to anticipate from that value. Therefore, we update $\boldsymbol{\theta}_t^u$, supposing the following equation

$$\boldsymbol{\theta}_{t+1}^{u}{}^{(j)} = a^* \boldsymbol{\theta}_t^{u(j)} + (1 - a^*) \boldsymbol{\theta}_{t+1}^{u*} \,, \tag{25}$$

where $\boldsymbol{\theta}_t^{u(j)}$ is $\boldsymbol{\theta}_t^u$ for the j^{th} particle. The first term relates to the case in which the desired position for user motion is the same as or close to the value at the previous time step. The second term relates to the case in which user motion changes suddenly, and the desired position for user motion is far from the value at the previous time step. a^* is the parameter that represents which term more strongly affects the value at the next step. $\boldsymbol{\theta}_{t+1}^{u*}$, which is not related to $\boldsymbol{\theta}_t^{u*}$, can be expressed by

$$\boldsymbol{\theta}_{t+1}^{u*} = \boldsymbol{\theta}_{t+1}^{u*}(\boldsymbol{D}_t^{\boldsymbol{y}}) \,, \tag{26}$$

where $\boldsymbol{D}_t^{\boldsymbol{y}} = \{\boldsymbol{D}_{t-1}^{\boldsymbol{y}}, \boldsymbol{y}_t\}$ is the information set at time t with regard to observation $\boldsymbol{y}_t$. Even in this case, (13) is still valid (with $\boldsymbol{\theta}$ replaced by $\boldsymbol{\theta}_{t+1}$). Therefore, comparing (25) with (19) (and (14)), it is expected that the same methodology can be applied by replacing $\boldsymbol{m}_t^{(j)}$ by $\boldsymbol{\theta}_{t+1}^{u}{}^{(j)}$ given in (25). The two cases differ in that $\boldsymbol{\theta}_t^{(j)}$ in (19) is the main part in (19) (a is large), while $\boldsymbol{\theta}_{t+1}^{u*}$ is the main part in (25) (a^* is small). The difference appears in the variance matrix. Over-dispersion of the variance matrix must then be checked carefully. In the estimation of the desired position for user motion, the variance matrix can be expressed as

$$(h^2 + a^{*2}) \boldsymbol{V}_t^u + (1 - a^*)^2 \boldsymbol{V}_t^{u*} \,, \tag{27}$$

where

$$\boldsymbol{V}_t^u = \sum_{j=1}^{N} \omega_t^{(j)} \left(\boldsymbol{\theta}_t^{u(j)} - \overline{\boldsymbol{\theta}_t^u}\right)\left(\boldsymbol{\theta}_t^{u(j)} - \overline{\boldsymbol{\theta}_t^u}\right)^T \,, \tag{28}$$

$$\boldsymbol{V}_t^{u*} = \sum_{j=1}^{N} \omega_t^{(j)} \left(\boldsymbol{\theta}_t^{u*} - \overline{\boldsymbol{\theta}_t^u}\right)\left(\boldsymbol{\theta}_t^{u*} - \overline{\boldsymbol{\theta}_t^u}\right)^T . \tag{29}$$

When $\boldsymbol{\theta}_t^{u*} \cong \overline{\boldsymbol{\theta}_t^u}$, letting $h^2 + a^{*2} = 1$ or $h^2 + a^{*2} < 1$, the variance matrix becomes $(h^2 + a^{*2}) \boldsymbol{V}_t^u$ (constant or decreasing), and the diffusion of dispersion can be avoided similarly to the original APF with parameter estimation. When $\left|\boldsymbol{\theta}_t^{u*} - \overline{\boldsymbol{\theta}_t^u}\right|$ is large, letting a^* be small such that the sampled particles can be around $\boldsymbol{\theta}_t^{u*}$, the variance matrix has small magnitudes (because the second term in (25) is the same for every particle), and diffusion can be avoided. In any case, with small a^*, diffusion can be avoided. In practice, if user motion changes suddenly, a small a^* is used. If user motion is not large, a^* satisfying $h^2 + a^{*2} = 1$ is used and the (constant) desired position for user motion can be obtained.

Summarizing, for estimating desired user motion position, we will use the algorithm given in Section III.C, replacing $\boldsymbol{m}_t^{(j)}$ by $\boldsymbol{\theta}_{t+1}^{u}{}^{(j)}$ given in (25) and controlling the magnitude of a^* such that the diffusion of the variance matrix given in (27) can be avoided.

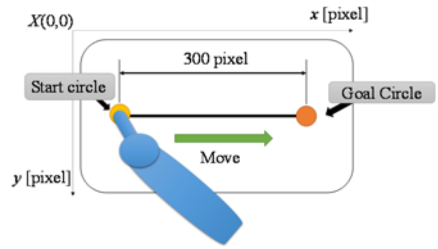


Figure 3. Overview on touchscreen for Experiment 1

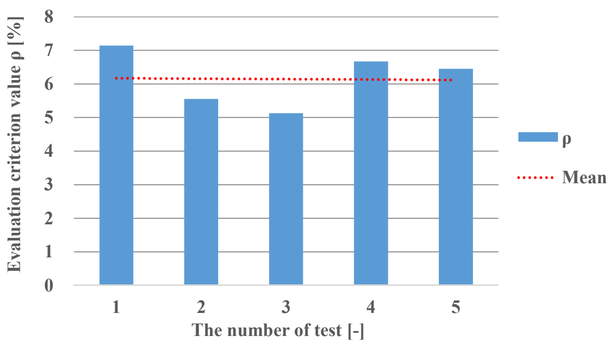


Figure 4. Evaluation criterion value ρ for every test in Experiment 1

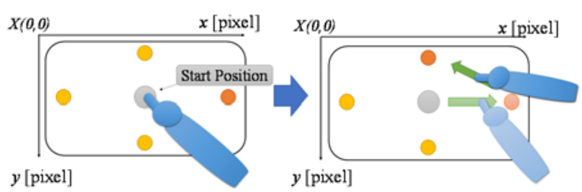


Figure 5. Schematic view of task in Experiment 2

V. EXPERIMENTS

The purpose of the presented algorithm is improvement of usability by predicting the desired position for user motion and displaying it. Three experiments were conducted to evaluate the validity of the algorithm.

A. Experiment 1

In the first experiment, we see whether the presented algorithm can actually anticipate user motion, predicting the desired position several time steps later.

Fig. 3 shows a schematic view of Experiment 1. The operator was asked first to put a fingertip on the start/yellow circle and move to the goal/orange circle without decreasing the speed (this means that the fingertip goes through the goal circle). The goal circle indicates the desired position of user motion in this case. If the anticipated state variable can reach the goal earlier than the actual fingertip gets there, then the presented algorithm can anticipate the user motion (intention). The time it takes to move from yellow/start to orange/goal circles was then investigated.

Sensors attached to the touchscreen are used for obtaining the position $\boldsymbol{y}_t$ of the user fingertip. Based on this observation value, the state and observation equations without noise terms are defined for user motion (intention) anticipation as follows:

$$\dot{\boldsymbol{x}}_t = \boldsymbol{K}_t(\boldsymbol{\theta}_t^u - \boldsymbol{x}_t) \tag{28}$$

$$\boldsymbol{y}_t = \boldsymbol{x}_t \tag{29}$$

where $\boldsymbol{K}_t$ is the gain and $\boldsymbol{\theta}_t^u$ is the desired position for user motion. Note that $\boldsymbol{\theta}_t^u$ corresponds to the estimated desired position for user motion, the user intention. As mentioned above, we investigated how far in advance $\boldsymbol{\theta}_t^u$ reaches the final goal compared to $\boldsymbol{y}_t$ (measured fingertip position). If $\boldsymbol{\theta}_t^u$ can arrive at the final goal earlier than $\boldsymbol{y}_t$, then the presented algorithm can display/predict user intention in advance of the user motion.

The ratio of the times for $\boldsymbol{\theta}_t^u$ and $\boldsymbol{y}_t$ to reach the goal gives us a quantified evaluation criterion and shows how stably the presented system can show the user intention in advance of the user motion. We define the criterion as follows:

$$\rho = \left(1 - \frac{t_r(\boldsymbol{\theta}_t^u)}{t_r(\boldsymbol{y}_t)}\right) \times 100 \ \ [\%], \tag{30}$$

$$t_r(\boldsymbol{p}) = t_e(\boldsymbol{p}) - t_s\,, \tag{31}$$

where

$t_e(\boldsymbol{p})$: The time/moment when $|\boldsymbol{p} - \boldsymbol{p}_g| < \varepsilon$ is satisfied, where $\boldsymbol{p}_g$ is the position of the goal circle and ε (= 10 pixels) is a small positive constant

t_s: The time when user motion starts

We used a Sony Duo 11 tablet PC with touchscreen (CPU: Intel(R) Core(TM) i7-3687U 2.10GHz, OS: Windows 8 Pro, Touchscreen size: 1920 × 1080 pixels, electrostatic capacity type touchscreen), while the screen size of the application was 900 × 600 pixels. The distance between the start and goal positions was set to 300 pixels. The sampling time was set to 10 msec. The number of particles was 100. The experiment was conducted five times.

Fig. 4 shows the results representing the criterion value at each test and the mean. In every case, a positive value can be obtained, and it can be seen that $\boldsymbol{\theta}_t^u$ can always precede $\boldsymbol{y}_t$; the factor was over 5% in every case, and the mean reached 6%. This shows that the system can display user intention (desired position) stably in advance of the user motion. Note that it does not matter whether the magnitude of 5~6% is large or not. What the magnitude is positive is important. These results show success in the anticipation of user motion.

B. Experiment 2

In the second experiment, we see whether the presented algorithm can actually deal with user motion that lacks the Markov property, a sudden change in user motion.

Fig. 5 shows a schematic view of Experiment 2. On the touchscreen, there are four circles, three yellow and one red. Initially, the user finger is positioned at the center of the screen. Then, the user was asked to move her/his fingertip to the red circle. If the user fingertip gets to the red circle, one of the yellow circles changes to red. Which circle goes to red was determined randomly. Then, we can observe the behavior of the system when user intention suddenly changes. The user tracked the red circle ten times. The other settings were the same as in Experiment 1.

Fig. 6 shows a portion of the results, because the rest of the results showed the same tendency, and space is limited. The blue and green lines show the observation values; the blue line shows the *x* coordinate of the user fingertip, y_x, while the green line shows the *y* coordinate, y_y. The red and yellow lines show the anticipated user position/intention; the red line shows the *x* coordinate of the position, θ^u_{xt}, while the yellow line shows the *y* coordinate, θ^u_{yt}.

It can be seen that the system can stably predict the desired position for user motion even when user motion changes suddenly. The successful anticipation of user motion can also be seen. The anticipated position θ^u_{xt}, θ^u_{yt} always precedes the measured position y_x, y_y, especially for high-speed motion.

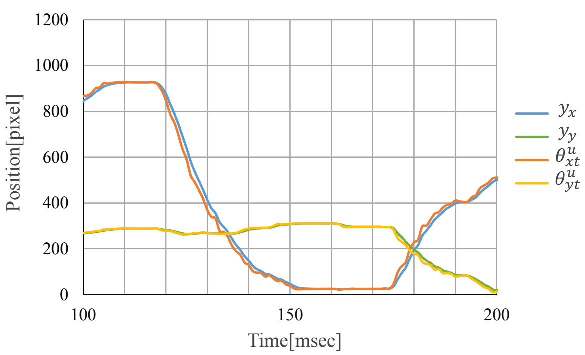


Figure 6. Time series data of measured and anticipated user motion $(\boldsymbol{y}, \boldsymbol{\theta}^u)$ in experiment 2

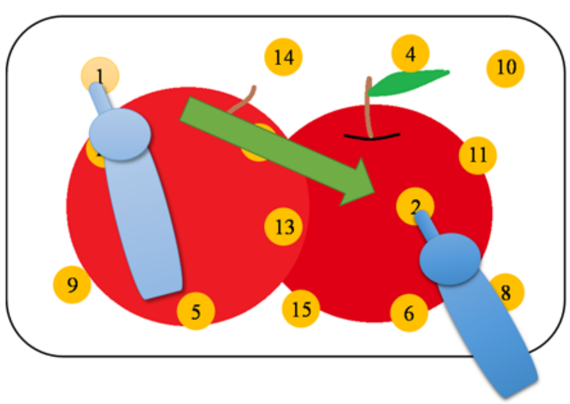


Figure 7. Schematic view of the game screen in experiment 3

C. Experiment 3

In the third experiment, we see whether anticipation by the presented algorithm can actually improve usability. For this purpose, a sensory examination method was conducted using a simple game application. Games were prepared with and without the presented algorithm, and we asked subjects to compare and evaluate the usability of the games.

Fig. 7 shows a schematic view of the game screen (1200 × 700 pixels) that the subject sees in the experiment. The numbered blue circles (from one to 15) are located randomly on the screen. The user is asked to wipe out all the circles with one fingertip touching the screen, maintaining touch throughout the game. If the fingertip is positioned around the circle, the circle disappears. However, the user has to wipe out the circles in ascending order. The following three kinds of systems were prepared for evaluation.

With both anticipation and sign: The line segment $\boldsymbol{S}_{\boldsymbol{y}-\boldsymbol{\theta}^u}$ from $\boldsymbol{y}$ to $\boldsymbol{\theta}^u$ was displayed as a sign, where $\boldsymbol{y}$ is the observed fingertip position and $\boldsymbol{\theta}^u$ is the anticipated desired fingertip position determined by the presented algorithm. The condition of the erasing circle is $|\boldsymbol{S}_{\boldsymbol{y}-\boldsymbol{\theta}^u} - \boldsymbol{p}_{ci}| < 40$ pixels, where $\boldsymbol{p}_{ci}$ is the *i*th circle position.

With only anticipation: The presented algorithm was used, but no sign was displayed to help users. The erasing condition is $|\boldsymbol{S}_{\boldsymbol{y}-\boldsymbol{\theta}^u} - \boldsymbol{p}_{ci}| < 40$ pixels.

Without anticipation or sign: The presented algorithm was not used. No sign was displayed to help users. The erasing condition is $|\boldsymbol{y} - \boldsymbol{p}_{ci}| < 40$ pixels.

The sampling time was 100 msec. The other settings were the same as in Experiments 1 and 2. We asked 20 subjects (22-24 years) to conduct the games with the three kinds of systems (three games per person), and to select one game/system as having the best usability. Fig. 8 shows that the presented algorithm improved usability, because the number selecting the system with the presented algorithm is greater than the number selecting the system without the algorithm. Whether a sign should be displayed is considered to depend on individual preference, as was confirmed using a written questionnaire.

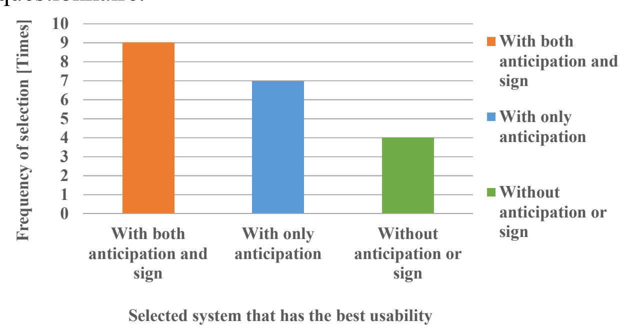


Figure 8. Frequency of selecting the game/system with the best usability

VI. Conclusion and Future Work

This paper presented a methodology for anticipating user motion on a touchscreen, aimed at improving usability. In order to have simple structure, only regular sensors originally attached to a touchscreen are used. User motion on a touchscreen can change suddenly, with consequent inability of the system to react to user motion. To overcome this problem and improve usability, this paper presented a

 ISBN: 978-1-61208-325-4

methodology for anticipating user motion several time steps later, based on APF with parameter estimation. Three experiments were conducted to evaluate the effectiveness of the presented approach. In the first experiment, we showed that the desired position of user motion (user motion several time step later) can be anticipated in every case. In the second experiment, we showed that the system can predict user motion even when user motion suddenly changes. The third experiment indicated that the anticipation of user motion/intention by the proposed algorithm can improve usability. These experimental results showed the validity of our approach. It is considered that there are many other systems whose efficiency can be improved by incorporating the presented algorithm. This will be the direction of our future work.

Acknowledgment

This work was partly supported by JSPS KAKENHI Grant Number 24650092 and the Telecommunications Advancement Foundation.

References

[1] Z. Chen, F. Lin, H. Liu, Y. Liu, W-Y. Ma, and L. Wenyin, " User Intention Modeling in Web Applications Using Data Mining," World Wide Web, 2002, pp. 181-191.

[2] M. G. Armentano and A. A. Amandi, "Recognition of User Intentions for Interface Agents with Variable Order Markov Models," User Modeling, Adaptation, and Personalization, 2009, pp. 173-184.

[3] M. Carrasco and X. Clady, "Prediction of user's intention based on the hand motion recognition," Proc. of Int. Conf. of the Chilean Computer Society (SCCC), 2009.

[4] A. Doucet, N. D. Freitas, and N. Gordon, "Sequential Monte Carlo Methods in Practice," Springer, 2001.

[5] M. Z. Islam, C. M. Oh, and C-W. Lee, "Video Based Moving Object Tracking by Particle Filter," Int. J. of Signal Processing, Image Processing and Pattern, vol. 2, no. 1, March, 2009, pp. 119-132

[6] Y. Wang and J. Wu, "Adaptive Particle Filter for Data Fusion of Multiple Cameras," J. of VLSI Signal Processing 49, 2007, pp. 363-76.

[7] K. Oka, Y. Sato, Y. Nakanishi, and H. Koike, "Head Pose Estimation System Based on Particle Filtering with Adaptive Diffusion Control," MVA, 2005, pp. 586-589.

[8] J. O. Wobbrock, M. R. Morris, and A. D. Wilson, "User-Defined Gestures for Surface Computing," Proc. of SIGCHI Conf. on Human Factors in Computer Systems, 2009, pp. 1083-1092.

[9] Y. Lin, "Evaluation of User Interface Satisfaction of Moblie Maps for Touch Screen Interfaces," Proc. of Int. Conf. on Advances in Computer-Human Interaction , 2012, pp. 22-27.

[10] K. C. Dohse, T. Dohse, J. D. Still, and D. J. Parkhurst, "Enhancing Multi-user Interaction with Multi-touch Tabletop Displays using Hand Tracking," Proc. of Int. Conf. on Advances in Computer-Human Interaction, 2008, pp. 297-302.

[11] M. Pitt and N. Shephard, "Filtering via simulation: Auxiliary Particle Filters," Journal of the American Statistical Association. vol. 94, no. 466, Jun, 1999, pp. 590-599.

[12] S. Thrun, W. Burgard, and D. Fox, Probabilistic Robotics. MIT Press, 2005.

[13] N. Vlassis, B. Terwiji, and B. Krose, "Auxiliary Particle Filter Robot Localization from High-Dimensional Sensor Observations," Proc. of IEEE Int. Conf. on Robotics and Automation, 2002, pp. 7-12, vol. 1.

[14] K. Goto, Y. Kawahara, T. Yairi, and K. Machida, "Anomaly detection for spacecraft by estimating parameters with particle filter," Trans. of the Japan Society for Aeronautical and Space Science vol. 55, no. 642, 2007, pp. 355-358.

[15] S. J. McKenna and H. Nait-Charif, "Tracking human motion using auxiliary particles filters and iterated likelihood weighting," Image and Vision Computing , vol. 25, 2007, pp. 852-862.

[16] M. West, "Approximating Posterior Distributions by Mixtures," Journal of Royal Statistical. Society(Ser. B), vol. 55, 1992, pp. 409-422.

[17] J. Liu and M. West, Combined parameter and state estimation in Simulation-based Filtering. Springer, 2001.

[18] N. J. Gordon, D. J. Salmon, and A. F. M.Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation ," IEE Proc. of F (Radar and signal processing), vol. 140, no. 2, 1993, pp. 107-113.

[19] M. Isard and A. Blake, "Condensation – conditional density propagation for visual tracking," Int. J. of Computer Vision, vol. 29, 1998, pp. 5-28.

[20] F. Dellaert, D. Fox, W. Burgard, and S. Thurm, "Monte carlo localization for mobile robots," Proc. of IEEE Int. Conf. on Robotics and Automation, 1999, pp. 1322-1328, vol. 2.

 ISBN: 978-1-61208-325-4

Panoramic Interaction with Interval Data Based on the Slider Metaphor

Phillip C. S. R. Kilgore, Marjan Trutschl, Urška Cvek
Department of Computer Science
Louisiana State University in Shreveport
Shreveport, USA
e-mail: {pkilgore, mtrutsch, ucvek}@lsus.edu

***Abstract*—Users utilize information visualization tools to explore their multivariate data in two- and three-dimensional environments in order to uncover patterns, form hypotheses or pursue questions of interest; they filter, search, zoom, and interact with the data. We first present a short survey of interactive selection mechanisms for information visualization and related work. We follow with our panoramic range slider and show the examples of its application and utility and present results of a case study. Finally, we elucidate on the advantages that the panoramic range slider can provide while identifying areas of further refinement.**

Keywords-range selection; range slider; linear control; interval data interaction; query inversion; visualization interaction; graphical user interfaces.

I. Introduction

With the exponential increase in scale and amount of data generated follows the need for tools for dynamic and intuitive interaction. Users want to be able to formulate their strategies, pose questions and uncover patterns of interest, to identify information hidden in such data. Information visualization is an effective means of making sense of multivariate data. We typically utilize the two-dimensional (2D) or three-dimensional (3D) space to plot data records as dots, bars, lines and glyphs, mapping dimensions to position, size, shape, color and, in some cases, motion. Interactivity with these data increases the utilization of the human visual perception, providing for a more meaningful analysis of the data [1].

Dynamic queries using interactive tools are a powerful means of data interaction, due to their visual nature and support for incremental, rapid visual actions on the data. They can elucidate global trends, and help find answers to specific questions [2]. These queries are formulated using widgets such as sliders, buttons and check boxes. Dynamic HomeFinder [3] was one of the first systems that successfully integrated dynamic queries into visualization.

A variety of techniques are utilized for exploratory analysis of data sets. Users commonly seek an answer to the question ”What is the typical behavior of the system from which we are collecting data?”. In such cases, the user is interested in data which is bound by some range specified in terms of a central tendency and variance. However, users sometimes look for data that deviates significantly from the norm (outliers).

An example of this can be seen during analysis of differential gene expression data, where interesting records are not representative of the majority of the data. In this kind of analysis, two or more microarray samples are taken with the intent of discovering genes which show significant changes in regulation, and the underlying hypothesis is that these genes belong to one or more biological pathways [4]. The expectation is that comparatively few genes will show regulatory changes, so outliers typically become objects of intense study.

While differential gene expression experiments are a good example of an outlier-based approach, there are other instances where this sort of analysis is warranted. For instance, an educational institution focused on improving standardized test scores (either globally or within some demographic) might wish to approach the task by either determining steps taken by the top-scoring students, which typically do not represent the majority of test-takers. For instance, in 2009, the average composite score for the American College Testing assessment was 21.1, while participants receiving the maximum score of 36 represented 0.04307% of all participants [5]. Since college admission and scholarship awards are often based on scores received on such tests, it is within these institutions' best interest to analyze the behavior of the outliers to claim notable improvements in their own population.

Finally, there exists the possibility of interaction with a measure that is intrinsically modular. Interaction with a closed dimension (a finite yet unbounded dimension such as degree measure or clock time) is an example of such a phenomenon. In these cases, we are dealing with half-open intervals of the form $[0, n)$, where $xn \equiv 0, x \in \mathbb{Z}$.

We present a brief overview of dynamic interactive queries, focusing on the scrollbar and slider metaphors. We discuss related work and present our panoramic range slider that can be utilized for selection in both closed and open intervals. Additionally, we show examples of its application using real-life multivariate datasets and conclude with a brief case study. Finally, we postulate several use cases for the panoramic range slider and discuss future research directions.

II. Related Work

Several continuous-domain controls in computing are metaphors for analog devices. For instance, the *slider* concept shares much in common with linear potentiometers (or ”faders”) and its availability is nearly ubiquitous. Sliders are typically implemented as discrete controls; however, with sufficient granularity, they can be used to emulate continuous value selection. Another popular metaphor is the *dial* concept, which is a metaphor for a radial potentiometer and is a core control in the Qt GUI framework [6]. *Thumbwheels* (or rollers) are another radial form of continuous input that (while less commonly implemented) can be found in various GUI toolkits [7]–[9]. AlphaSlider [10] is a query device that allows users to select one item from an alphabetically arranged list without using a keyboard.

These metaphors are useful for input of a single, scalar value. However, range input is a more difficult task to approach because there are few metaphors for which a tangible analog instrument exists. A commonly provided form of range-based input includes the *scrollbar* concept, which can provide input specifying a range with a fixed width. The *piano roll* metaphor commonly found in music production software (such as FL Studio [11], Reason [12], and Rosegarden [13]) is continuous with respect to time and is based off of the recording media used for player pianos. The piano roll also has an interesting property: its ability to effectively visualize the content of the underlying document. Daschelt et al. provide a method of facet-based navigation called *FacetZoom* [14] based on a stack of one-dimensional treemaps that is suitable for navigating (amongst other things) textual data.

A. Interval-Based Interaction

The above range-based controls show weaknesses when attempting to manipulate numeric intervals. The piano roll metaphor may be useful for time-domain visualization and manipulation; however, these are vector data, and numeric intervals tend to be a subset of a scalar domain. Scrollbars allow for the positioning of ranges, but do not intrinsically permit selection of range width. Finally, FacetZoom [14] works best with a finite subset of its underlying domain, as it was designed to expedite navigation to a single datum rather than a range of data.

A range slider (Fig. 1) is a control used to specify continuous intervals. On the left is a minimum bound indicator a (Fig. 1.a), and on the right, the maximum bound indicator c (Fig. 1.c). The space occupied by the combination of the bound indicators and the thumb represents a subinterval of the interval $[0, 1]$ (Fig. 1.d). Suppose that the bounds indicators have a width of w_b, that the thumb has a width w_t, and that the bounding area has a width w. Let x_a denote the position of the left side of a, and x_c denote the right side of c. Then, the range slider represents the closed interval $[x_a/w, x_c/w]$, and $w_t = x_c - (x_a + 2w_b)$.

Range sliders share much in common with the scrollbar metaphor, although a notable difference arises out of its usage. Like a scrollbar, a range slider consists of a thumb (or slider, Fig. 1.b) that is free to move within a rectangular bounding area (Fig. 1.d). Unlike a scrollbar, the two buttons (the bound indicators) associated with a range slider are connected to the thumb and define its width. Common applications of range sliders focus on filtering a range of records. Becker [15] used a double-edged slider with upper and lower thresholds on network maps, Eick et al. [16] used a categorical slider on software displays for selecting an arbitrary subset from hundreds of time-ordered software modifications, and Ahlberg [17] used a suite of double-edged sliders for FilmFinder. Spotfire [18] was the first commercially successful information visualization tool based on traditional dynamic queries utilizing range sliders.

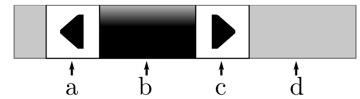


Figure 1. A range slider consists of several parts: a) minimum bound indicator, b) thumb, c) maximum bound indicator, and d) bounding region. The interval depicted is $[a, c]$.

The *pie of time* [19] is a circular range slider widget that allows selection of times of day. The inner circle has the full 24 hour cycle around its perimeter. The hours can be individually switched on or off and a range of hours can be clicked and dragged to enable/disable them. The *radial range selector* is a similar concept by Till Nagel [20], and gives the ability to select a range on any circular, repetitive event data, such as the time of day, weeks or seasons, and other circular data (such as angles, etc.).

Over the past couple of decades range sliders have been extended to spatial data to interactively select records within spatial proximity, such as the *TrapezoidBox* [21]. The TrapezoidBox interface allows the users to combine related query results using interactions. For medical data sets, Müller et al. utilized a *brushing histogram* view overlaid with a value range as a histogram, on which the user can make a selection [22]. A study of dynamic query sliders versus the brushing histograms was done by Li et al. [23] where they used both in an empirical experiment on a geographic data visualization tool. A user study led them to believe that the brushing histograms were superior for complex discovery tasks. A number of similar range-based selection mechanisms have been introduced, including parallel coordinate displays where each axis is a controlled attribute range, queries for temporal variations, peaks, valleys and slopes, and their selection includes not only the mouse and keyboard, but touch, gesture and speech. Readers are encouraged to visit [24] for a full taxonomy of interactive dynamic tools for visual analysis.

B. Problem Analysis

The range slider metaphor is applicable for both closed and open intervals. However, it may be useful to exclude a range, such that for all x, $x \notin [a, c]$, or equivalently, $x \in [0.0, a) \cup (c, 1.0]$. A naïve solution to this might involve linking a checkbox to the inverted interpretation of a traditional range slider, as seen in Fig. 2. When the control is used in a nested list of checkboxes, problems can occur. There, it is not clear what the checkbox affords: does the checkbox afford inversion, or (as an example) the domain's consideration for a dynamic query? It is also unclear what the checked state means even in the context where it is known to be associated with the slider.

This problem could (in part) be resolved by descriptive text provided by a label or tooltip; with a label, ambiguity is mitigated at the cost of increased usage of the display medium. Tooltips could also be used in this capacity (albeit with some introduction of delay) but this method does not lend itself well to environments which lack a mouse cursor (such as mobile devices). This control invites error, either in the form

 ISBN: 978-1-61208-325-4

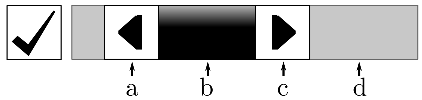


Figure 2. A possible incarnation of a range slider affording inversion by using a checkbox to toggle the inverted state. This method may not cascade well in nested lists involving other checkboxes.

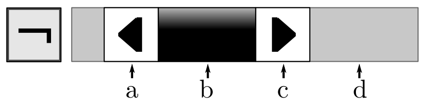


Figure 3. A slight improvement on the modal approach to range slider inversion is provided by replacing the checkbox with a push-button. The effect of the control is no longer ambiguous, but mode errors are still possible.

of misinterpretation of the checkbox's function or in the form of mode error.

Recognizing that the checkbox is being used in a modal capacity, a possible improvement might be to replace the checkbox with a push-button depicting a glyph that explicitly labels the mode (Fig. 3). The button remains depressed during inversion; otherwise, it is raised. If an appropriate glyph is chosen, there is no longer any question as to what the control affords, and the mode of the slider can be discerned from the state of the button. This method also happens to violate the expected behavior of push-buttons in some environments (where they are monostable controls and the push state is expected to be transient) [25]–[27]. Some user interface (UI) toolkits do provide for bistable push-button controls [6], where it might be acceptable for a push-button to maintain a depressed state; however, while it allegedly improves upon the initial checkbox design, the construction presented in Fig. 3 still leaves the possibility for mode error.

A better alternative to prevent mode error would be removing the mode controlling inversion to begin with. One such example of this is the radial range selector presented by Nagel [20]. There are no modes implied in its design, and although no mechanism for range inversion is presented, it can be afforded by reversing the bound indicators. This control was designed with interaction of closed dimensions in mind, and in principle, it could be used with open dimensions as well.

However, there exist a number of issues which make it less useful in the latter capacity. Because it is designed to interact with cyclic data, there is no indication where the interface between the bounds (l and h) of an open dimension is. While intuition might tell us that this position is at 12:00, this point could (in principle) be anywhere, including the 3:00 position associated with a traditional Cartesian plot of $(x, y)_\theta = (\cos\theta, \sin\theta)$. Without any indication of this position, it would lead us to believe that the notion of a start position is arbitrary, but this is not the case with an open dimension, as they have a precise lower and upper bound.

This issue might be solved by explicitly marking this interface (as depicted in Fig. 4), but that introduces another problem. The presence of this boundary implies (by reasoning analogous to the interpretation of a clock face) that the lower (l) and upper (h) bounds are equivalent; again, this is not

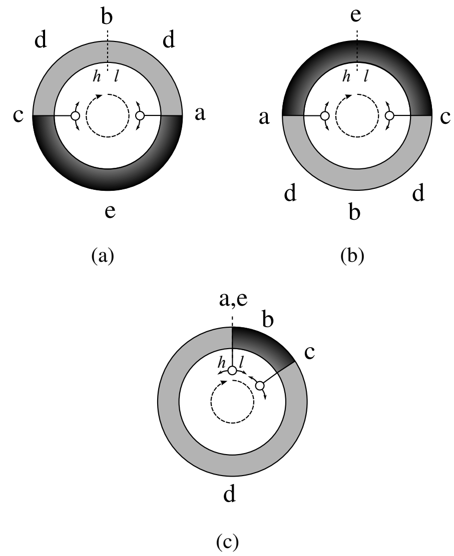


Figure 4. An augmented radial slider with a bound interface indicator. a) The interquartile range is selected. b) The inverse of (a), the extremes, is selected. c) A selection with an ambiguous lower bound because the domain boundaries (l, h) overlap.

so with an open dimension. While this affords an inverted range where including the extremes would be desirable, it also affords inconsistent behavior of that interface.

Consider the use case where a radial range slider is used to filter property values. Any interval within the interquartile range is unambiguous: knowing that the interface represents the absolute bounds, we assume that the selection opposite to it must represent some range in the middle of the absolute range. It might be useful for the user to query values within the interquartile range of property values; such a selection is easily recognized because the semicircle opposite to the bounds indicator must indicate this half of the range (Fig. 4.a). Likewise, it might be desirable to look at the extremes in the data by inverting the selection; again, the semicircle intersecting the bound interface succinctly depicts this (Fig. 4.b).

When the selection sits on the threshold of the bound indicator, difficulty in interpretation can arise (Fig. 4.c). Here, a selection of the lowest value is probably intended, but it is questionable whether or not the highest value is also selected, because it shares the interface with the lowest value. This is a consequence of the metaphor. For a closed dimension, this is the expected and proper behavior: after all, $0° \equiv 360°$ in degree measure, and $0{:}00 \equiv 24{:}00$ on the face of a clock. Thus, regardless of the direction of the approach, the limit at any point is preserved, because the two values are (by definition) the same.

This is not the case with an open dimension. Here, the interface represents a jump discontinuity where the limits from either side are not equal; the interface (in fact) represents values with maximal distance. A reasonable interpretation of Fig. 4.c might conditionally represent either extreme based on the limit of the opposite direction to the handle, but this now overloads the behavior of the interface, because intersecting the interface indicator will represent two values.

 ISBN: 978-1-61208-325-4

III. Panoramic Range Sliders

While the radial range slider has issues, it has several appealing properties. Firstly, the selection range (with the expectation of the discontinuity at the bound interface) is clear, and this reduces the need for a mode to specify inversion. The lack of a mode is also appealing. If we could indicate for the open dimension that there was a definite, non-equal bound, we would solve one of the primary issues with the radial range slider. Thus, we utilize the linear depiction of the range.

One question that arises is how to visually depict inversion of the selected range. This can be shown by having the arrows on a classic range slider point inwards and to provide a thumb on both extremes of the bounding area (Fig. 5). The area represented by d is left unshaded to indicate that it is not part of the selection. We name the semantic extension of this metaphor a "panoramic range slider" due to its similarity to $360°$ panoramic photographs.

If one glues together the left and and right edges of the bounding area, an annulus is formed. Let τ be any point in $\mathbb{R}$ along the circumference of the outer rim, such that $\tau = 0$ denotes the point where the bounding area was joined, and let τ_l and τ_h respectively represent the positions of the lower and upper bounds. The mapping to the interval $[0.0, 1.0)$ can be calculated by the bijection $\tau(\theta) = \theta/2\pi$.

As with the traditional range slider, the resulting range is a subset $\mathcal{P}$ of $[0.0, 1.0]$. However, the circles are themselves modular in nature, such that any value of θ can be mapped to $[0.0, 2\pi)$ via the binary operator "$x \operatorname{fmod} y$". The fmod operator is surjective and is defined by (1). Thus, $\mathcal{P}$ can be defined as the set specified by (2).

$$x \operatorname{fmod} y \equiv x - y\lfloor x/y \rfloor \quad (1)$$

$$\mathcal{P} \equiv \forall \theta \in (\mathbb{R} | x_l < \theta < x_l + x_h), \exists x \in [0, 1] : \tau(\theta \operatorname{fmod} 2\pi). \quad (2)$$

In this sense, the range slider as presented in Fig. 5 is somewhat like the ancient symbol Ouroboros, which depicts a snake eating its own tail. However, we find that if we straighten out this proverbial "snake" by ungluing the vertices at the new boundary, one may occasionally obtain two non-null subsets of $\mathcal{P}$: the set $[0, a)$ and the set $(c, 1]$. This happens when $x_a > x_c$, or when the minimum is greater than the maximum. We refer to this state as "negated state", such that the slider is interpreted to mean $x \notin [a, c]$.

It should be noted that the underlying dimension need not be intrinsically modular, just that it must be totally ordered and have definite bounds. By gluing the bounds together, the range slider is in principle the radial slider presented in Fig. 4. However, the linearization offers two new pieces of information. The range covered by the two handles is the range that will be selected; it is also apparent that there is a minimum and maximum in this range.

There are several ways that we suggest the slider could enter a negated state. One is to cross the bound indicators over another (e.g., moving the maximum indicator before the minimum indicator); another way is to move the thumb beyond the bounding region. Both of these methods have a severe drawback, as the user must carefully reposition the bound in an attempt to enter negated state, likely leaving an interval with different bounds and introducing error. Alternative mechanisms for triggering inversion are possible, however; in an environment for which a keyboard is present, the use of quasi-modal interaction (such as ctrl-click) can quickly and accurately result in the inversion. Other possible mechanisms include selection via an entry in a context menu.

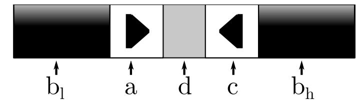


Figure 5. A panoramic range slider in negated state. The interval depicted is $[0.0, l) \cup (h, 1.0]$. A panoramic range slider is a linear realization of the radial range slider control in Fig. 4. Note that in its negated state, there are two handles (shown here as b_l and b_h).

IV. Evaluation

As a proof of concept, we have chosen to apply these techniques to a data set containing both open and closed dimensions. We generated a synthetic data set containing 10,000 randomly-generated samples of the tuple $(x \in \mathbb{Z}, y \in \mathbb{Z}, z \in \mathbb{Z}, r \in \mathbb{R}, \theta \in \mathbb{R})$. Each point is constrained within a maximum radius R and maximum height Z, such that $0 \leq r \leq R$ and $0 \leq z \leq Z$. Thus, a cylinder of radius R and height Z is a convex set encapsulating the data, and the points have Euclidean coordinates $(x = \lfloor r \cos \theta \rfloor, \lfloor r \sin \theta \rfloor, z)$. It should be noted that the data is devoid of semantics, and the task at hand is the selection of regions within that convex set.

The range sliders contribute to a 5-volume (the bounding volume) which consists of the bounds of four open dimensions and the closed dimension θ (Fig. 6). Intervals for r, θ and z can be combined to form a bounding volume specified by an annular sector prism. It is reasonable that r and z can be specified as open dimensions without any loss in usability, since these dimensions do not require modularity. A traditional range slider provides a proper subset of possible bounding prisms because it can hand intervals for θ such that $\theta_l \leq \theta_h$. However, when $\theta_l > \theta_h$, two bounding prisms must be employed: one which that represents the interval $[0, \theta_h]$ and one that represents the interval $[\theta_l, 1.0]$.

The lack of modular behavior in the traditional range slider can interfere with the effective specification of the endpoints of the bounding prism. For instance, the disjoint nature of the two bounding prisms means that it is possible to accidentally specify two separate bounding prisms rather than the two intersecting ones. Secondly, in a complex data query, we are usually interested in the intersection of records which fit each constraint in the data query. However, a special exception to the semantics of the data query must be made to treat the two range sliders as a union and causes some difficulty in understanding the semantics of the query.

An alternative method is to duplicate each record, and to substitute θ with θ_{+360} in the duplicate. This method doubles the range of θ and therefore permits a single range slider to encapsulate the selected range; this results in data

 ISBN: 978-1-61208-325-4

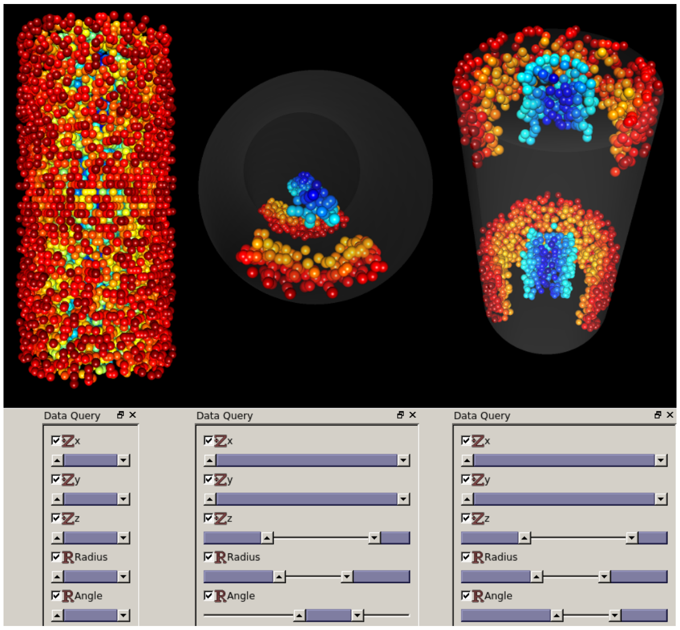


Figure 6. A collage of various selections for the above data set (records colored by radius). The second and third selections cannot be expressed by a single ordinary range slider per dimension.

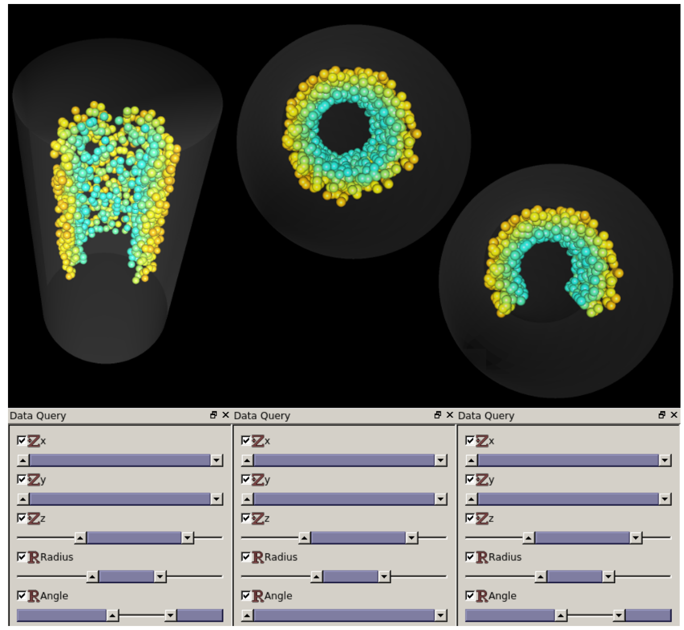


Figure 7. Various selection bounds on the data presented in Fig. 6 using only the three range sliders provided in the data query. The second angle dimension is no longer necessary and can be omitted from the data entirely.

redundancy, which may require infeasible amounts of CPU time and memory for large data sets.

Panoramic range sliders address this problem because the underlying abstraction involves a closed dimension. The query is simplified because the disjunction is not directly exposed to the user. A single panoramic range slider produces a single set which has exactly two bounds to be manipulated by the user (Fig. 6, Fig. 7). Thus, while it is reasonable to view the resulting set as the union $[0, \theta_h] \cup [\theta_l, 1]$, the modular nature of the control also means that the user views the set as a single interval $[\theta_l, \theta_h + 2\pi]$ when $\tau_l > \tau_h$.

Fig. 6 demonstrates the behavior of the panoramic range slider. The first selection captures the entire data set. The second selection demonstrates both traditional range slider behavior (applied to the angle) and sliders with negated state. The third selection is similar, but it negates the angle. Neither of these are afforded by the ordinary range slider, which requires the interval to be contiguous. Accurate inversion is afforded by a hotkey (Fig. 7).

In both Fig. 6 and Fig. 7, neither the dimension θ_{+360} nor data duplication is necessary when using a panoramic range slider. Because θ is inherently modular, it fits the concept of the panoramic range slider extremely well. Furthermore, the complement of the bounding volume is easily obtained by inverting both the z, r, and θ components (Fig. 6).

V. Conclusions and Future Work

The addition of modularity to range sliders provides greater control over how data is filtered in several ways. First, panoramic range sliders permit the concept of excluding data that falls within a given interval in addition to the inclusive approach taken by traditional range sliders. Secondly, it allows inherently closed dimensions to be represented in a way that reflects their modular nature. Finally, it preserves bounds, making it simpler to determine what the effective bounds represented by the data are.

The data set used during the evaluation has served as a viable proving ground for these properties for two reasons. The data set is sufficiently large: in the case of the provided data set, a high degree of occlusion can be seen such that interior records are difficult to investigate. The data set is also dense: in this data set, it is uncommon to find a record without a neighbor close by.

Several panoramic range sliders in concert provide the ability to create complex bounding volumes. In that case, we were able to create several bounding solids permitting us to perform a variety of queries based on three dimensions alone. Complementary solids can also be formed, since the range slider permits inversion.

In the future, we intend to perform a detailed user study focusing on user interaction with panoramic range sliders. In addition to comparing the modular range slider to its traditional variant, we will compare it to other continuous, interval-based input controls. Finally, we intend to investigate methods of further improving the panoramic range slider by including widgets such as context menus to provide exact input or lock the behavior of the panoramic range slider.

Acknowledgements

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Numbers P20GM103424 and 8P20GM103433. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

[1] R. Spence, Information Visualization: Design for Interaction, 2nd ed. Prentice Hall, 2007.

[2] B. Shneiderman, "Dynamic queries for visual information seeking," IEEE Softw., vol. 11, no. 6, Nov. 1994, pp. 70–77, [retrieved: Jan., 2014]. [Online]. Available: http://dx.doi.org/10.1109/52.329404

[3] C. Williamson and B. Shneiderman, "The dynamic homefinder: evaluating dynamic queries in a real-estate information exploration system," in Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval, ser. SIGIR '92. New York, NY, USA: ACM, 1992, pp. 338–346. [Online]. Available: http://doi.acm.org/10.1145/133160.133216

[4] M. Kerr and G. Churchill, "Experimental design for gene expression microarrays," Biostatistics, vol. 2, 2001, pp. 183-201.

[5] "ACT National and State Scores," 2011. [Online]. Available: http://www.act.org/newsroom/data/

[6] Digia, Plc., "Qt project," [retrieved: Nov., 2013]. [Online]. Available: http://qt-project.org

[7] B. Spitzak et al., "Fast Light Toolkit," [retrieved: Nov., 2013]. [Online]. Available: http://www.fltk.org/

[8] Silicon Graphics Inc., "IRIX 6.5 documentation," 1998, man page: `SgThumbWheel(3)`.

[9] Kitware, Inc., "KWWidgets," [retrieved: Nov., 2013]. [Online]. Available: http://www.kwwidgets.org/

[10] M. Osada, H. Liao, and B. Shneiderman, "Alphaslider: Searching textual lists with sliders," in Proceedings the 9th Annual Japanese Conference on Human Interface. ACM Press, 1993, pp. 365–371.

[11] Image-Line Software, "Fl studio 11," [retrieved: Jan., 2014]. [Online]. Available: http://www.flstudio.com

[12] Propellerheads AB, "Reason," [retrieved: Jan., 2014]. [Online]. Available: http://www.propellerheads.se/products/reason

[13] C. Cannam et al., "Rosegarden: music software for Linux," [retrieved: Jan., 2014]. [Online]. Available: http://www.rosegardenmusic.com

[14] R. Dachselt, M. Frisch, and M. Weiland, "Facetzoom: a continuous multi-scale widget for navigating hierarchical metadata," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '08. New York, NY, USA: ACM, 2008, pp. 1353–1356. [Online]. Available: http://doi.acm.org/10.1145/1357054.1357265

[15] R. A. Becker, S. G. Eick, E. O. Miller, and A. R. Wilks, "Dynamic graphical analysis of network data," in ISI Conference Proceedings, 1989, pp. 119–120.

[16] S. G. Eick, J. L. Steffen, and E. E. Sumner, Jr., "Seesoft-a tool for visualizing line oriented software statistics," IEEE Trans. Softw. Eng., vol. 18, no. 11, Nov. 1992, pp. 957–968, [retrieved: Jan., 2014]. [Online]. Available: http://dx.doi.org/10.1109/32.177365

[17] C. Ahlberg and B. Shneiderman, "Visual information seeking: tight coupling of dynamic query filters with starfield displays," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '94. New York, NY, USA: ACM, 1994, pp. 313–317, [retrieved: Jan., 2014]. [Online]. Available: http://doi.acm.org/10.1145/191666.191775

[18] C. Ahlberg, "Spotfire: an information exploration environment," SIGMOD Rec., vol. 25, no. 4, Dec. 1996, pp. 25–29, [retrieved: Jan., 2014]. [Online]. Available: http://doi.acm.org/10.1145/245882.245893

[19] Stamen Design, "Oakland Crimespotting update: the pie of time," [retrieved: Nov., 2013]. [Online]. Available: http://content.stamen.com/node/171

[20] T. Nagel, "A radial range slider," [retrieved: Nov., 2013]. [Online]. Available: http://www.tillnagel.com/2011/02/radialrangeslider

[21] M. Cho, B. Kim, D. K. Jeong, Y.-G. Shin, and J. Seo, "Dynamic query interface for spatial proximity query with degree-of-interest varied by distance to query point," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '10. New York, NY, USA: ACM, 2010, pp. 693–702, [retrieved: Jan., 2014]. [Online]. Available: http://doi.acm.org/10.1145/1753326.1753428

[22] H. Mueller, K. Zatloukal, M. Streit, and D. Schmalstieg, "Interactive exploration of medical data sets," in Proceedings of the 5th International Conference on Information Visualization in Medical and Biomedical Informatics, 2008, pp. 29–35.

[23] Q. Li, X. Bao, C. Song, J. Zhang, and C. North, "Dynamic query sliders vs. brushing histograms," in CHI '03 Extended Abstracts on Human Factors in Computing Systems, ser. CHI EA '03. New York, NY, USA: ACM, 2003, pp. 834–835, [retrieved: Jan., 2014]. [Online]. Available: http://doi.acm.org/10.1145/765891.766020

[24] J. Heer and B. Shneiderman, "Interactive dynamics for visual analysis," ACM Queue, vol. 10, no. 30, 2012, pp. 30–55.

[25] User Experience Interaction Guidelines for Windows 7 and Windows Vista. Microsoft Corp., Sept. 2010.

[26] OS X Human Interface Guidelines. Apple Inc., Oct. 2013, [retrieved: Jan., 2014]. [Online]. Available: https://developer.gnome.org/hig-book/3.0/

[27] GNOME Human Interface Guidelines 2.2.2. The GNOME Project, 2013, [retrieved: Jan., 2014]. [Online]. Available: https://developer.gnome.org/hig-book/3.0/

Electronic Health Records and the challenge to Master the Patients' Pathways. Proposals around a comparative analysis (France / Spain).

Biedma José Maria
Business Organization Department
University of Cadix
Cadix, Spain
e-mail: josemaria.biedma@uca.es

Bourret Christian
DICEN IDF Research Team
University of Paris East
Marne-la-Vallée, France
e-mail: bourret@u-pem.fr

Abstract - **Issues about patients' pathways and interoperability of Information Systems are at the heart of the challenges faced by Healthcare Systems to control costs and also to improve quality of care. We propose a comparative analysis of the issue of computerized patients' records in France and Spain highlighting requirements, similarities and differences and new possibilities of actions. Most of the challenges faced by the Healthcare Systems converge on this issue. After an analysis of the broader context of Healthcare Systems in these two countries, which have different designs, but have also similar problems, we'll examine the issue of the Healthcare record ownership, then the access to its data and its handling of doctors' personal notes. We'll also discuss ethics' issues and especially that of medical confidentiality. We'll then consider the central topic of interoperability linked to different approaches: national in France and at the level of the Autonomous Communities in Spain. The changes are only at their beginning. The use of these new socio-digital tools necessarily meets the European dimension, the issue of using open data and that of digital territories in Healthcare.**

Keywords - Electronic Health Records ; Healthcare ; patients' pathways ; access ; interoperability.

I. INTRODUCTION

All the Healthcare Systems in developed countries, particularly in the European Union, are in crisis, facing many problems, both to control costs and to improve quality of care. The development of socio-technical devices within the broad context of Information and Communication Technologies (ICT) induces deep changes and provides new opportunities. The solutions are sought through the patient's pathway, i.e., traceability issues with information and communication challenges. Shared Electronic Health Records can be considered as "hologrammatic" (when a part contains the same properties as the whole system cf. E. Morin and J.-L. Le Moigne ideas on Intelligence of the complexity [1]) of the whole Healthcare Systems' challenges. If the problems are the same everywhere, the national or regional approaches may vary. In this paper, which is a joint work between two researchers (Spanish and French), we propose the outlines of a comparative analysis of the requirements, the challenges faced in the two countries with proposals for improvement and opportunities for future research studies.

After an introduction, we will show that the two Healthcare systems have the same requirements but in different contexts. Then we will analyze some specific aspects of the Electronic Health Records in the two countries: ownership, access, the collection of data with the issue of physician's personal notes. Afterwards, we will discuss the question of the interoperability, ethics' aspects and especially medical confidentiality. After regarding towards a European record and the question of open data, as a conclusion, we will outline that we are only at the beginning of profound changes.

II. SIMILAR REQUIREMENTS IN DIFFERENT CONTEXTS

Spain and France have a different approach of Healthcare issues. In France, the governance of the Healthcare System is traditionally centralized, but with a distribution of roles. The State controls the system (regulation) and the supervision of the Healthcare Organizations. Since 1945, the different Health Insurance Offices: *Caisse Nationale d'Assurance Maladie des Travailleurs Salariés* (CNAMTS), MSA (*Mutualité Sociale Agricole*) and RSI (*Régime Social des Indépendants*), are supervised on a parity way by the social partners (both employers and trade unions representatives). They manage the funds on the basis of contributions coming both from employees and employers. It is a so-called Bismarck's system [2] (from the name of the German Chancellor who created this type of Social Insurance System in 1875). In France, since the 1980s, we also get an important part of resources coming from taxation: General Social Contribution (CSG). Since the mid-1990s, the system has a more regionalized management, with the recent creation of the Regional Health Agencies (ARS) in 2010, but managed at the regional level by State officers (*déconcentration*) and not by Autonomous Regions (*décentralisation*). In Spain, the Autonomous Communities appeared with the Status of Autonomy (1978). They have gradually gained the control of the management of their citizens' Health. They have developed specific policies. This was particularly the case for Catalonia (1981), for the Basque Country (1982) and Andalusia (1984). These Regional Healthcare Systems are financed through taxation: Beveridge's systems [2] (from the name of the designer of the British Health National System in 1946, mainly the NHS or National Health Service).

There is a major difference between Spain and France. In Spain, the Communities are autonomous and develop specific tools and may have specific legislation for Health data, in addition to those of the Spanish State, especially in Catalonia. In France, the main tools, including the Personal Health Record (DMP: *Dossier Médical Personnel*) of the patients, whose construction began in 2004, are implemented from a national perspective. In Spain, they are in a large part developed at the regional level, which immediately raises the question of the transfer of data and systems' interoperability.

As all the other developed countries (United Kingdom, Germany, United States of America, etc.), France and Spain have quite similar requirements to improve different aspects of their Healthcare Systems.

First, they must master the growing costs: 11.7 % of the GDP (Gross Domestic Product) in France and 9.4 % in Spain in 2011. They must also improve the quality of care with a key idea: avoid the breaks in the traceability of care, which is costly and does not improve patients' care. Consequently, the central issue has become that of patients' pathways. Thus, it is imperative to master the use of new tools, especially the Electronic Health Records (EHR).

Other important requirements are both security of managed information and that of interoperability between various technical devices.

If the requirements are roughly similar, the general contexts are different: national in France and rather regional in Spain, as we have outlined above.

The two systems are also different in the way the medical profession is exercised and remunerated in primary care. In Spain, as in the United Kingdom, the General Practitioners (GPs) often work in cooperative organizations and receive a global amount for each registered patient. In France, the GPs' activities is private and still essentially isolated: each GP has his own office. The French GPs are not globally financed but paid for each accomplished medical act by the patients, the patients being reimbursed later by the different Health Insurance Organizations.

In France, the EHR's issue is particularly sore. The GPs fear that new electronic tools may be tools to control their activity and, consequently, their incomes. So, in 2004, the DMP has been named "personal" and not "shared". It has been decided that it will be hosted by private companies.

The question of the ownership of the EHR is essential.

III. THE OWNERSHIP OF THE PERSONAL HEALTH RECORDS

We have already mentioned that in France, the patient's medical record was (Health Insurance Law of August 2004) designed as "personal" and not "shared". For some people, the weight of the medical lobby was very heavy in Parliament to avoid controlling the activity of Physicians by Health Insurance Companies. The patient owns his personal medical record.

But, since 2004, things have gradually evolved towards a more "shared" DMP. First, medical practice is a little less individual and new cooperative organizations have formed (Healthcare Networks, Homecare, medical centers, etc.).

Then the two public consortia managing the DMP, the GIP- DMP and, after 2008, the ASIP (Agency for Shared Information Systems in Health) have had significant problems. In December 2013, only 500,000 records were opened, compared to 5,000,000 contractually hoped for, and very few are really operational. With this scandal, some people talk to entrust the management of DMP to CNAMTS but then the fear that the DMP then becomes primarily a monitoring tool for rationalization of the activities of the doctors and not only and principally to improve patients' pathways reappears .

As a result of his ownership, the patient may "hide", even destroy some data. Some spoke about a "naked doctor in front of a masked patient." What is the medical value of a non-exhaustive record? The issue is all the more serious since in France the responsibility of the physician remains individual. The DMP is still experimental although originally its generalization was scheduled for 2007.

In Spain, Law 41/ 2002 of November 14 does not tackle this issue which is still debated. We can distinguish four positions. One group believes that the ownership of personal medical records must belong to the health center ([3], [4] and [5]). For the second group ([6] and [7]), the property of the personal medical records must belong to the doctor. This position's basis lies in the notion of the doctor's copyright. This does not mean that the patient doesn't have a right of access its content. A third group, as Gay Montalvo, believes that the ownership of personal medical records must belong to the patient [8]. For other authors [9], the question of the ownership of personal health records is a problem that is irrelevant and of little practical importance. What is really important is to know who gets the right of access.

The debate has rebounded with differences in the various Autonomous Communities. Some (Galicia and Valencia) consider that the property belongs to the government or to a health center if the doctor works for others. If the medical profession is exercised individually, the property belongs to the physician. Other Autonomous Communities have not yet legislated on this subject.

IV. ACCESS TO ELECTRONIC HEALTH RECORDS

The issue of access (consents) is also a key one. As a result of his ownership of the record, in France, the patient gives the access rights (authorizations) to the DMP. He must have access to hospitalizations' data and copies of his hospital records. This follows from the Law of 4 March 2002 on the Rights of Sick People and the Quality of the Health System, confirmed by a law of August 2004 on the evolution of Health Insurance (providing experimentation for DMP) and different decrees in 2006. In case of emergency, when patient's life is engaged, the physician can "force" the access, but the situation is not satisfactory. There are two opposite camps: the “fundamentalists” of the absolute property of the patient and those who insist that the DMP is not a safe for personal data but simply a tool to improve patients’ care and traceability. Recent developments seem to go to a shared record [10].

In Spain, Law 41/ 2002 (Article 18) extended the right of access the patient's personal medical records and he is allowed to get copies of the data. Health centers must comply with this patient’s right, exercised directly by them or by a duly accredited representative.

The right of access to medical records is not an absolute right. According to art. 18.3 of Law 41 /2002, the exercise of this right has two limitations: first it cannot be exercised to the detriment of others' right to confidentiality of the information it contains and the therapeutic benefit in relation to the data collected from the patient. It may be exercised at the expense of professionals who have participated in the preparation of this medical story. Yet, there is a limit derived from the rights of the professionals involved in the development of medical records, which may restrict the right of access, excluding its subjective annotations made in the record [11].

Another situation is the access to a record belonging to a dead person, or to a mentally ill person or a minor one. The law allows health centers and individual exercise practitioners to provide access to medical records of patients who died or persons associated with them for family or factual reasons (art.18.4 Ley 41/2002). However, this is denied if the dead people had expressly forbidden it. Some authors point out that the right of access to personal medical records is only for the patients and is not transferable to their heirs. Therefore, in order to get them it is necessary to prove the cause and the legitimate interest of the applicants in the context of a judicial or administrative proceeding [12].

The question of the age of the child to exercise his right of access to personal medical record by himself is also problematic. Law 41 /2002 does not address this issue explicitly. As recalled by Saiz and Larios [13], the doctrine has been applied by analogy to the minimum age (sixteen years, unless emancipation), according to Article 9 of the Act on the Protection of Data for the provision of informed consent without representation. However, for these authors, the entire solution is currently governed by Article 13 of the Regulations of the Organic Law on Data Protection that allows a fourteen years old person to gain access to their personal data "except where the Act requires assistance for holders of parental authority or guardianship." It is therefore necessary to adopt a uniform approach to determine the age when the child can get the right to this access.

V. DATA COLLECTION FOR PATIENTS’ RECORDS: THE ISSUE OF PERSONAL PHYSICIANS’ NOTES

After the questions of the ownership of the patient’s record and of its access, another one is that of the data it contains.

These questions are raised in different ways in France and Spain, according to different specific approaches.

In France, the central issue of the collect of data, patients’ data or not, is a main one. Since 1927, the physician has been paid directly by the patient for each medical act. The patient is then reimbursed by the various Health Insurance companies. This is the origin of so-called French “liberal system”. Private primary care physicians did not want to be paid directly by the Health Insurance organizations which could thus control their activity. The central issue (often not formulated officially), already outlined, is always the fear of Health Insurance organizations’ control of doctors’ activity and of their incomes. Many blockages around the DMP come from this not formulated but very real fear. So who will fill the DMP? If the data is not validated by the doctor it has no medical value. The doctor did not incur individual responsibility for a medical record that would have been filled by the patient. In times of severe budget pressures, the Health Insurance Organizations do not want that the collection of patients’ data (and also the reconstitution of their personal health history in the past) could become a new and therefore reimbursed medical act. The problem underlines a limit of the doctor’s payment for each medical act [14].

The question can be expressed otherwise: that of the "double collection" of patient’s data by a physician. With the developments in ICT, some optimists believe that a single “click” will switch patients’ personal records with data coming from the personal physician's file concerning the same patient. But then comes the main question of the personal physician’s notes, their ownership and their confidentiality.

The approach is different in Spain. The concept of subjective notes or annotations of the doctor gives way to multiple interpretations. Lorenzo (2006) [15], according to Cantero [5], considers that the notes are personal impressions of the doctor on the attitudes of patients who were tested objectively. For Sanchez-Caro and Abellán [16], subjective notes include comments or personal impressions of the doctor. If they do not have clinical importance, they should not be included in the clinical history. On this issue, the Spanish Justice emphasized that these notes are "assumptions made on prints that do not strictly correspond to the content of the record" (STSJ de Canarias 48904). It is also understood that they are "personal impressions of the patient or his social environment, attitude or behavior of the patient's reactions (STSJ Madrid, 2006/162708). The physician can be attacked on personal notes which were not intended to circulate.

Other authors have objected to the exclusion of subjective notes. For Galán [17], if these possibilities of entries are deleted, they eliminate 90% of the clinical history of the patient. The author argues that these terms should be maintained, because deleting them would act against the fundamental right of patients to access to a document. This would have a negative effect in terms of an investigation or assistance. Other authors believe that the support for this limitation is in accordance with the right to privacy [13].

Regional laws that have tackled the issue of subjective annotations are rare. We can give the example of Article 32 of Law 3/2005 of 8 July (Health Information and patient autonomy) in Extremadura. This article examines the subjective annotations that are considered as "impressions of

 ISBN: 978-1-61208-325-4

health care based on the unique perspective of people, and that in any case are not relevant to the true knowledge and update status patient's health, but cannot be considered a diagnosis." Note also Article 19 of law 3/ 2001 of 28 May the Government of Galicia, which concerns informed consent and medical history of patients. According to Troncoso [18], the right of access to personal medical records cannot affect the rights of health care professionals. Therefore, they may refuse to provide access to their subjective annotations.

Saiz and Larios [13] point out that, in theory, the patient has the right to know and obtain copies of all reports and documents listed in Article 15 of Law 41 /2002. However, in practice, it is not uncommon to find some medical institutions whose internal rules only allow specific delivery of data and diagnostic reports. This is also the case in France where patients do not get the copies of their documents as easily as the law requires it.

A solution would be the standardization of sizes and types of documents that must be provided to patients and to establish greater control of this aspect of the patients' rights to information in health organizations, in order to avoid situations like those listed. In this perspective raises another question: it is not the same thing to write notes for yourself or to be read by others (question of semantic interoperability)?

VI. The Central Issue of Interoperability

Interoperability is not considered in the same way in France and Spain due to different approaches: national (France) and regional (Spain).

In Spain, the lack of standardization of Personal Electronic Health Records creates problems when the patient moves from one Autonomous Community to another one, with a loss of a chance of survival in case of disease or accident. The launch of the Online Health is a progress. Its purpose is the exchange of clinical information between Autonomous Communities with the "Digital Health Record" project of the Spanish National Health Service. The second phase of the implementation will be completed in 2014. On 20 September 2013, the Council of Ministers has approved the establishment of a National Health Card in the next five years to replace all other cards [19].

For Criado [20], the project is not yet achieved due to the difficulties to reach an agreement between all the actors of the system, despite the fact that the majority of semantic, technical, organizational and even interoperability problems have been resolved. There are 16.7 million electronic medical records existing, but four autonomous communities have not yet joined the project [21].

The Spanish Health System is moving towards a shared and interoperable computerized clinical history. As in France, it concerns three types of professionals: computer scientists, healthcare professionals, clinical documentation workers. To improve the interoperability of the Electronic Health Record, it is necessary to incorporate reference models that also allow to correctly identify the contents [22].

A very effective element to support this interoperability effort is the use of EN / UNE 13606 (Health informatics - Electronic health record communication) standard [23]. But, in addition to the use of a reference model (standard), Spain must gradually develop archetypes, a decisive technical artefact to incorporate clinical knowledge in existing electronic documents from the design phase of the implementation of the system. Since the project "*Historia Clínica Digital del Sistema Nacional de Salud* (HCDSNS)", significant progress has taken place in the construction and dissemination of archetypes in a large scale, they give support to the requirements of the National Health System projects.

According to the Study of Information Technology and Applied Communication for Health and Inclusion, carried out in 2011 [24], a fundamental challenge for Health Information Systems is interoperability. This means exchanging clinical information so that patients can have their clinical data available in any place. As this study demonstrates, from a technological point of view, interoperability needs Enterprise Application Integration (EAI), Enterprise Service Bus (ESB) or Service Oriented Architecture (SOA) platforms.

Benson [25] considers that the development of interoperability in the health field needs the use of standards. These include:

- SPL HL7 (Structured Product Labeling): electronic labeling standard for drugs.
- HL7 [26] Medical Records: for the management of medical records.
- HL7 Clinical Genomics: This standard facilitates the exchange of customized clinical data between multiple agents: assistance providers, laboratories and centers for biomedical research.
- DICOM [27] (Digital Imaging and Communication in Medicine). It is a worldwide recognized standard for handling, storing, printing and transmitting medical images between different information systems.

Other resources are devoted to the development of projects for the security of information systems and the implementation of technological platforms for the integration of systems that enable interoperability of Electronic Health Records in each region.

In France, a major issue is that of the divisions (*cloisonnements*): between the Ministry of Health and the various Health Insurance organizations, between primary care and hospital sector, between public and private sectors, between doctors and other healthcare professionals, between care sector and the social one, etc.

The question of boundaries may also be involved in the great number of various patients' medical records. In France, the highlighted DMP is nevertheless far from unique [28, 29]. Patients may have different records: DMP, DCC (Communicating Record for Cancer), Drug History (*dossier pharmaceutique*), records of patients in different care or medical-social organizations, and also records of different interface structures (Networked Organizations, Health centers, hospitalization at home, etc.).

 ISBN: 978-1-61208-325-4

Would not all these personal patients' records, expected to help to break down barriers in the Health sector, create new ones finally? A great change is necessary because their multiplication is the source of not essential expenses and of confusing problems. But there are also many competing interests.

The authorities seem to be aware of the problem. During the presentation of the National Health Strategy on 2013 September 23th, the Minister of Health announced the "progressive construction of a Public Information Service in Health" and a wider opening of health data [30]. Initially, it should be designed to validate sites offering health information and ultimately to coordinate all institutions providing and managing health data. The unfortunate experience of DMP still at the experimental level and the financial situation may contribute to temper the initial enthusiasm.

It is also necessary to overpass an initial confusion. The DMP has become a computerized patient record among others. We need a coordination tool for interoperability at national level: the DMP or another device [31].

The issue of interoperability does not have only a technical aspect for cooperation between different socio-technical systems but also a semantic aspect, in terms of vocabulary between different professions (engineers and professionals of care), concepts and national cultures [32]. So if in Spain and France national starting points are different, the problems converge around the notion of interoperability and the need for a tool for national coordination.

VII. Ethics' aspects and medical confidentiality

In the health field, the use of the technologies of information and communication (ICT), is now irreversible and has undeniable advantages, such as traceability of actions (care pathways), but, at the same time, may raise fears. Some, like Romeo and Castellano [3] fear an increased risk of violations of privacy and other individual interests. Therefore the French DMP arises ethics' issues. Three different points emerge very quickly: the issue of the chosen security devices, the use of personal data in relation to the specificity of the health data: their medical confidentiality.

The security issue is primarily that of the systems used. Many examples are discussed showing loss of data, errors in transmission of electronic documents or in data anonymization. Cañas and Santander [33] stress the risk of intentional attacks (malicious) or that of employees' error, which may affect the confidentiality of information transmitted or manipulated.

But we must not exaggerate. The digital data are more largely protected than the paper patient records lying around on tables or in cabinets in medical offices or in hospitals.

Then comes the question of personal data. Nowadays, each country has its own approach resulting from various collective mentalities.

Anglo-Saxon countries do not have the same approach as the Latin countries. The United Kingdom, always very concerned by the civil liberties, is repugnant to the development of national cards, like identity or health cards. In the United States, freedom of use of personal data is much more important, and so the risks of skidding are higher. The Congress instructed the Health and Human Services Department (HHS) to protect the patients' privacy by integrating its aspects into the Health Insurance Portability and Accountability Act (HIPAA) of 1996 (Office for Civil Liberties). Each federate State can add specific aspects. In Quebec (Canada), the Commission of Access to Information has a specific role.

Spain has an agency for data protection (*Agencia Española de Protección de Dados*). Under the Statute of Autonomy of 1978, some autonomous communities have established their own organizations to protect the rights of their citizens. In Catalonia, an *Agència Catalana de Protecció de Dades* (APDCAT) was been established by the Organic Law 15/1999 on Data Protection. They also have an *Administraci ó Oberta de Catalunya* which aims at developing a "digital citizenship". We can find too an *Agencia Vasca de Protección de Datos* (AVPD). The *Agencia de Protección de Datos* of the Madrid's Community was closed in 2013.

Citizens' attitudes towards personal data (privacy) may vary. A survey conducted in the United States on adults by Harris Interactive in March 2003 [34] distinguish the "privacy fundamentalists" (26%), the "privacy pragmatists" (64% compared to 54 % in 1994) and the "privacy unconcerned" (10 % compared to 22% in 1994). Nearly 70 % of surveyed adults agreed with the statement that "consumers have lost all control on information collected and used by companies" and the majority are relatively pessimistic about the impact of legislation, 53% do not agree on the fact that "existing laws and organizational practices provide a reasonable level of protection for the consumer today".

There exists fears with the development of Big Data (large and unstructured data that invade networks and storage systems of all kinds) uses and, above all, about the risk of sales of patients' data to organizations for commercial purposes, for example of patient's personal data to pharmaceutical companies, without patients' consent.

The European Directive of October 1995 extended the protection of individual rights, introducing the concept of personal data, understood as any data that allows the identification of an individual. The United Kingdom adapted its legislation with EU Directive in October 1998. France will put a lot more time. Begun in 2002, the legislative process has been only finalized by the Act of 6th August 2004.

On 21th October 2013, the European Parliament voted a major revision of the 1995 Directive to strengthen citizens' control over their personal data. In the following years, a new EU regulation is going to replace the various national regulations. If it succeeds, in Europe, national approaches should be progressively erased.

Health data are personal and sensitive data but with strong specificities especially regarding medical confidentiality.

The concept of medical confidentiality is very old. It goes back to the famous Hippocratic Oath (from the name of the famous Greek physician of Antiquity), taken by all physicians who specifies that, "admitted in the privacy of people, I will not reveal the secrets that will be entrusted to me" [35].

For Dr. P. Cressard: "Medical confidentiality is the sum of the exchanges collected in a context of trust between a doctor and a patient" [36].

In France, the concept of breaking medical confidentiality appeared in 1810 under the First Empire. It has been extended to other professions in 1992.

The Law of 4th March 2002 on "Sick people's rights and the quality of the Health System [37] introduces major changes: medical confidentiality has become a right of the patient and not just an ethics' obligation of the physician. It placed the patients at the center of all decisions that affect them, introducing the concept of "health democracy." This law has been adapted to the hospital sector in particular with the 2nd March 2006 circular on the "Charter of the hospitalized person and the rights of users."

For Pr Anne Laude, the current state of the law reflects the partitioning (divisions) of the French Healthccare system: it would be preferable that a single text should bring together all the citizen's rights, whatever the nature of the medical act to simplify the work and the responsibility of physicians [36].

Thus, in France, but it can be generalized, the concept of medical confidentiality is going to evolve: with the impact of the new technologies of information and communication (ICT) at the heart of this article, but also because of the evolution of the practices of medicine. The medical practice is less isolated. Interface Organizations between isolated primary care and hospital have developed: Healthcare Networks, Homecare Organizations (in French: *Hospitalisation à Domicile* - HAD), medical houses, etc. Many actors are involved in the patients' care. But the responsibility of the physician and of medical confidentiality remains individual. What information is the doctor permitted to share?

The term of trust is often used in the relationship between doctor and patient or in the use of socio-technical devices (ICT). The "digital trust" is an essential issue. For Kaplan and Francou [38], we may invent new tools to rebuild the relationship between organizations and individuals and new creative spaces to develop trust. In France, perhaps the new interface organizations in Healthcare can play this role.

VIII. PROSPECTS: TOWARDS A EUROPEAN HEALTH RECORD AND THE CHALLENGE OF OPEN DATA

Today, in the European Union, health issues are always managed by the different national States. But we have seen that for personal data, national legislations had to adapt to the European Directive of 1995 and that they should gradually give way to a single European regulation. Moreover, since 2002 we have got a European Health Insurance Card, which permits medical or hospital care be obtained in a foreign country without advance payment. This is particularly important in the tourism sector or for people who move a lot [39].

Various projects have been developed in the European area to improve interoperability in the Health sector. This is for example the Calliope (European thematic Network for eHealth Interoperability) [40] project, HITCH (Healthcare Interoperability Testing and Conformance Harmonisation) [41], and especially, the epSOS or European patients Smart Open Services project [42]. It is the largest project in the European e-Health sector which emphasizes interoperability. The epSOS project is of great interest to facilitate the evolution of the national personal records to a European record.

Its main goal is to improve the health monitoring of citizens when they are outside their origin country, allowing health professionals of any country participating to the epSOS' project access to his medical data. To achieve this, providers of health services participating to the project cooperate by sharing their data to test this new service. For the first time, European people can use these cross-border services when they need health services in all countries participating in epSOS as a tourist, businessmen, students, travelers or tourists. This project is still in progress. In Spain, are associated the Ministry of Health, Social Services and Equality and five Autonomous Communities are associated: Andalucia, Castilla la Mancha, Catalonia, the Balearic Islands and Valencia.

In France, the project is managed by the ASIP [Agency for Shared Information System in Health) and is especially tested in the Ile-de-France region.

For Criado [20], the role of the European Union in this area is primarily for general guidance and incentives. We have shown important first steps for interoperability of data, imperative for data exchange in all the European Union territory.

Beyond the national projects, a more ambitious project would be the creation of a European personal medical record for each European citizen, which would allow access to his record from any country of the European Union. Of course the Health's sector is still a national competence, but the European Union is trying to facilitate cooperation to converge to a European health policy. There is already a European Health Insurance Card. It is only an administrative card and must be renewed each year. But it could be an entry for the access of the patient's data, including first emergency data. It would be an important step to a European e-Health.

Another point to outline is the use of digital data to improve care at a local level. In France, at the end of 2013, a new specific project was launched around areas of digital health or territories of digital care (in French, TSN: *territoires de soins numériques*) to develop the uses of digital data to improve the patient's care on specific territories. In an other manner, couldn't we also consider the creation of a specific new digital territory at the level of each patient representing a specific and personal digital care territory around its own personal medical record? This issue is very connected with telemedicine. Its development depends on specific national regulations (in France, 19th October 2010 decree).

Another important point is the challenge of the Open Data. In France, the CNIL, already mentioned, organized a seminar about this issue last July [43]. Participants stressed the idea that the open data is primarily a tool for modernizing public action and of general interest for democracy which needs the best opacity and privacy for citizens and transparency of the State, while it is often the opposite that happens. Open Data is an essential component of transparency. Personal data is not *a priori* the first concerned by the Open Data. Open Data have no personal nature, they concern mainly State data (maps, raw data, general indicators).

But we have entered a period of "revolution of the data." "Putting the world in data" lead us to redefine the way we act, and to create knowledge infrastructure of a new kind. The boundaries between the different categories of data, personal or public, anonymous or indirectly identifiable, are not always clear. "To build an Open Data legitimate and sustainable, it is essential to examine ethically the role of personal data, anonymisation, consent, etc."

IX. CONCLUSION: ONLY THE BEGINNING OF PROFOUND CHANGES

Privacy, ethics, consent, and also the concepts of access, interoperability, ownership; we find all the essential points of this article on the issues surrounding the Electronic Health Records and the challenge to master the patients' pathways through proposals around a comparative analysis in France and Spain.

The question of the implementation of electronic medical records is at the heart of the challenge of mastering the patients' pathways, both in France and Spain. We are only at the beginning of great changes caused by the rapid development of e-Health. In the future, the EHR that nowadays concern national or regional levels, will also be effective at local level and European level.

According to E. Morin's idea, it is also largely "hologrammatic" (representing a summary of various problematics) and at different levels.

Firstly, the EHR is "hologrammatic" for the challenges tackled by the two Healthcare Systems, and more largely those of all developed countries, facing the double problem of controlling costs and improving quality of care, in a particularly hard financial situation.

New ICT tools can help. The issue of electronic patient data is at the heart of our evolving health systems with the central issues of traceability of patients' pathways in care Patients become involved in their health and co-producers of new services concerning themselves, data and experiential knowledge such as tools, interoperability and the development of new uses and care practices. This is one aspect of the concept of "health democracy" (*démocratie sanitaire*) proclaimed by the Act of March 4th, 2002 which also insists on the idea of patients' responsibility, the Anglo-Saxon speaking of "empowerment".

But, secondly, it is also "hologrammatic" in a wide perspective around the "data revolution" with the definition of different data (public, personal, sensitive …), the questions of personal data and open data. So it is a global societal subject.

The newspaper *La Tribune.fr* (January 10th 2014) points out that "France is struggling to bring its Health System in the Digital Age". For the authors of the article, with the affirmation of the patient 2.0, the DMP reviewed, essential for the coordination of care, "may be the passport to the world of e-health" [44].

REFERENCES

[1] E. Morin and J.L. Le Moigne, The Complexity's Intelligence, in French : L'intelligence de la complexité, Paris, L'Harmattan, 2003.

[2] M. Duriez and D. Lequet-Slama, The Health Systems in Europe, In French : Les systèmes de santé en Europe, Paris, PUF, 1998.

[3] C.M. Romero and M. Castellano, "Patient privacy from the perspective of medical confidentiality and access to medical records", in Spanish: "La intimidad del paciente desde la perspectiva del secreto médico y del acceso a la historia clínica", D y S, 1993, vol. 1, 1, pp.5- 10.

[4] J.M. Álvarez-Cienfuegos, "Clinical history: custody and property (II) ", Presentation at the First Conference of Data Protection of the Community of Madrid", in Spanish: "La historia clínica: custodia y propiedad (II)", Fundación Mapfre-Medicina y Agencia de Protección de Datos de la Comunidad de Madrid, Madrid, 2000, pp. 123-128.

[5] R. Cantero, "Questions relating to medical history (I) and (II)", in Spanish: "Cuestiones relativas a la historia clínica (I) y (II)", La Ley,1996, pp. 1424 - 1430.

[6] A. De Lorenzo, "Who owns the medical records?", in Spanish "¿De quién son propiedad las historias clínicas?", Deontología, Derecho y Medicina, Ilustre Colegio Oficial de Médicos de Madrid, Madrid, 1977, pp. 491 - 495.

[7] A. Luna and E. Osuna, "Medico-legal storage and custody problems for the medical records" in Spanish "Problemas médico-legales en el almacenamiento y custodia de la historia clínica", Medicina Clínica, 1987, vol. 118, 2, pp. 631-635.

[8] E. Gay, "The patient's right to health and to clinical documentation. Problem on the property of the medical record", in Spanish: "El derecho a la salud y documentación clínica del enfermo. Problema sobre la propiedad de la historia clínica y su caso", Todo Hospital, 1987, 42, pp. 31-36.

[9] L. Sarrato, "The legal regime on access to medical records and guarantees", in Spanish: "El régimen legal de acceso a la historia clínica y sus garantías", Revista Jurídica de Castilla y León, 2009, 17, p.185-190.

[10] C. Bourret, "Issues of Information and Communication for New Managerial approaches: the Challenges of the French Health System", Journal of Computing and Information Technology –

CIT, University of Zagreb (Croatia), 2012, vol. 20, n° 3, p. 195 – 200.

[11] A. Garriga, Processing of Personal Data and Fundamental Rights, in Spanish: Tratamiento de Datos Personales y Derechos Fundamentales, Ed. Dykinson, Madrid, 2004, p. 119.

[12] J.M Méjica, The history, basic statute and proposed regulations, in Spanish: La historia clínica, estatuto básico y propuesta de regulación, Edisofer, Madrid, 2002.

[13] M. Saiz and D. Larios, "The right of access to medical records by patients: proposal for reservation of subjective annotations", in Spanish:"El derecho de acceso a la historia clínica por el paciente: propuesta para la reserva de anotaciones subjetivas", Derecho y Salud, 2009, Vol. 18, 1, pp.30, 36.

[14] C. Bourret,"Electronic Health Record (Dossier Médical Personnel) as a major tool for Improving Healthcare in France: an approach through the Situational Semiotic", The Second International Conference on Networked Digital Technologies (NDT), Prague, July 7-9, 2010, in F. Zavoral et al. (Eds.): Proceedings NDT 2010, Part II, Communications in Computer and Information Science (CCIS) Series of Springer LNCS, 88, pp. 7 – 14.

[15] O. Lorenzo, "Problem of subjective annotations in the 41/2002 Law", in Spanish: "Problemática de las anotaciones subjetivas de la Ley 41/2002", Paper XII, Congreso de la Asociación Española de Derecho sanitario, 2006.

[16] J. Sanchez-Caro and F. Abellan, Right and duties of patients, in Spanish:Derecho y deberes de los pacientes, Madrid, Ed. Comares, 2003.

[17] J.C. Galan, "The allusion to subjective annotations in the Autonomy Law creates uncertainty", in Spanish: "La alusión a las anotaciones subjetivas en la Ley de Autonomía crea inseguridad" Diario Médico, 2006, p. 10.

[18] A. Troncoso, Data protection in search of equilibrium, In Spanish: La protección de datos en busca del equilibrio, Tirant lo Blanch, Valencia, 2010, p.1186.

[19] Available on : http://www.elmundo.es/elmundosalud/2013/09/20/noticias/1379685783.html [Access Dec. 2013].

[20] J.I. Criado, "Interoperability and health policy in Spain", in Spanish : "Interoperabilidad y política sanitaria en España", Barataria, Revista Castellano-Manchega de Ciencias Sociales, 2013, 15, pp. 85 - 90.

[21] Available on : http://www.abc.es/sociedad/20130618/abci-historia-clinica-digital-201306181702.html [Access Dec. 2013].

[22] A.Romero, "Project HCDSNS on Medical Records in the National Health System, Progress and Prospects", in Spanish: "Proyecto HCDSNS Historia Clínica Digital del Sistema Nacional de Salud Avances y Perspectivas", Papeles Médicos, 2012, vol. 21, pp. 8-12.

[23] Available on : http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=50122. [Access Dec. 2013].

[24] Study of the Information and Communication Technologies. Applied to Health and Inclusion. Study of Art:e-Health and e-inclusion, in Spanish: Estudio de las technologias de la Informacion y de la Communicacion Aplicadas a la Salud y a la Inclusion, 2011, pp. 105-110.

[25] T. Benson, Principles of Health Interoperability, New York, Springer, 2010.

[26] Available on : https://www.hl7.org/ [Access Dec. 2013].

[27] Available on: http://www.dicom.com/fr/index.php [Access Dec. 2013].

[28] M. Fieschi, "To personal medical records. Patient data shared: an asset not to spoil to change the healthcare system" , in French : « Vers le dossier médical personnel. Les données du patient partagées : un atout à ne pas gâcher pour faire évoluer le système de santé », Droit social, 2005, n° 1, pp. 80 – 90.

[29] M. Fieschi dir., " Patient's shared data: the culture of sharing and of quality information to improve the quality of care", in French, « Les données du patient partagées : la culture du partage et de la qualité des informations pour améliorer la qualité des soins », 2003.

[30] Available on : http://www.acteurspublics.com/2013/09/23/vers-un-service-public-d-information-de-sante [Access Dec.2013].

[31] T. Ficatier, "Information System to coordinate care: performance driver " , in French : "Système d'information de la coordination des soins : levier de performance", Revue hospitalière de France, 2008, n° 521, p. 30-33.

[32] M. Fieschi, Governance of semantic interoperability is at the heart of the development of health information systems, in French : La gouvernance de l'interopérabilité sémantique est au cœur du développement des systèmes d'information en santé, Paris, 2009.

[33] F. Canas and F. Santander, "New technologies and confidentiality", in Spanish: "Nuevas tecnologías y confidencialidad", Asociación Española de Derecho Farmacéutico (A.S.E.D.E.F.), 2002, p 29.

[34] Available on : http://www.harrisinteractive.com/harris_poll/index.asp?PID=365) [Access Dec. 2013].

[35] Conseil National de l'Ordre des Médecins, Code of Medical Ethics, in French : Code de déontologie médicale, Paris, Jouve, 2004, p. 63.

[36] "Medical confidentiality between patients' right and ethical obligation", in French: " Le secret médical entre droit des patients et obligation déontologique" , Special issue nov.-dec. 2012, Bulletin d'information de l'Ordre national des médecins, p. 5.

[37] Law of 4 March 2002 on patients' rights and the quality of the health system, in French : Loi du 4 mars 2002 relative aux Droits des malades et à la qualité du système de santé.

[38] D. Kaplan and R. Francou, The Digital trust, in French : La confiance numérique, Paris, Ed. Fyp, 2012.

[39] C. Bourret, "Tourism, Leisure and Health: the issue of cross-border cooperation in case of accidents", in Spanish: "Turismo, Ocio y Salud: la cuestión de la cooperación transfronteriza en caso de accidentes", 1er Congrés d'Oci I Turisme, Mataró – Barcelona, May 2012.

[40] CALLIOPE Project, available on : http://esante.gouv.fr/en/node/45 [Access Dec. 2013].

[41] HITCH Project, available on: http://hitch-project.eu/sites/www.hitch-project.eu/files/HITCH-Roadmap-V1.1.pdf. [Access Dec. 2013].

[42] EpSOS Project, available on: http://esante.gouv.fr/le-mag-numero-3/le-projet-europeen-epsos [Access Dec. 2013].

[43] CNIL (National Commission on Informatics and Freedom), Seminar "Open data, which issues for data protection? "in French : Séminaire « Open data, quels enjeux pour la protection des données ? », Paris, July 9th, Available on : http://www.cnil.fr/fileadmin/documents/approfondir/dossier/OpenData/CR_Workshop_Open_Data_9_juillet_2013.pdf. [Access Dec. 2013].

[44] Available on : http://www.lesiss.org/445_p_37398/en-2014-a-votre-e-s@nte.html. [Access Dec. 2013].

Resocialization and Metaphor of Social Networks

Co-construct with the user relevance of relational technologies

Treton Cécile
Paris Est University
Paris, France
e-mail: Cecile.treton@paris.fr

Bourret Christian
DICEN IDF/-CNAM – UPEM (Paris Est University Marne-la-Vallée)
Paris, France
e-mail: bourret@u-pem.fr

***Abstract*— The subject of this publication focuses on the isolation problem. This can be considered a health and social problem since it has adverse health consequences to seniors. The objective of this research is to identify the relational patterns of the elderly in order to help recreate a network of relationships suited to their unique needs. The research method questions relationship dynamics, which include interactions with for example, alliances and desalliances, common points and affinities, meeting places including the use of social networks on the web. We study the new relational technologies around the notion of situation, described by Alex Mucchielli. The proposed work will rely on observation methods and interviews with isolated elderly. We want create a model of intervention which uses narrative methods for understand how the person built its social network.**

Keywords- ICT; social networks; isolated elderly; situation; frames; life stories; empathy; relational technologies

I. INTRODUCTION

The heat wave of 2003 and its consequences, 15 000 deaths, revealed in France as well as in Europe the problem of the loneliness of the elderly in our urban societies.

Social networks based on Information and communications technologies (ICT) can perhaps offer the means to rebuild a new social relation with the seniors, their entourage and their caregivers (actual or potential) to fight against loneliness.

This is the central hypothesis of a work in progress and the major axis of a doctoral thesis that we would like to present in this paper. Our theoretical context of research uses Alex Mucchielli [1] theory about “situational semiotic” and the principles of “engaging communication” described by Françoise Bernard [2]. The text is composed of five parts: position and method, definition of the problem, and axis of research declined through three chapters: empathy, networks and socialization and forms of socialization.

II. POSITION AND METHOD

This work is in the field of Science of Information and Communication (SIC), as proposed by Françoise Bernard [2], which revolves around questions of meaning, relationships, knowledge and action. We have adopted a constructivist approach with reference to Jean-Louis Le Moigne [3] that considers social reality as constructed by the actors involved, with questions about social representations and interactions.

We propose an approach of resocialization of isolated people, particularly the elderly, with the new relational technologies around the notion of situation, describe by Alex Mucchielli [1] like “situational and interactional semiotics”. Motivation for the action of an individual is built around different settings and contexts. We use narratives methods to understand how the elderly conceive friendly relationship. In order to facilitate the analysis, we will design a board with various criteria to define. A. Mucchielli proposed breaking down this situation, for background interpretation, into “frames”.

These frameworks are defined intentions and perceived stakes by actors, culture and standards of reference (values and representations), their positioning in relation to other actors, the quality of relationships maintained, historical and current framework, sensory part. Previously, all different actors have been identified.

Jean Jouet [4] explains that ICT change the social relationships and question of the planned, desired ICT uses while questioning the logic of diversion or unforeseen uses.

Our position of researchers is that defined by Nathalie Heinich [5] "commitment by neutrality": "engage in neutrality, produce action - not just knowledge - by highlighting coherence, logic, links, which, beyond the conflicts, make possible on at least a dialogue, and perhaps result in an inventive acceptable compromise "(analytical-descriptive perception). This work is also aimed at professionals and volunteers. The principles of “engaging communication” described by Françoise Bernard [2] can help us to describe a method for changing perceptions.

III. ISOLATION, A SOCIAL PROBLEM

A study of the Fondation de France [6], points out a decline in the integrative capacity of family, friends and neighborhood networks. It focuses on increasing situations of loneliness, which affect the entire population, especially the elderly. Between 2010 and 2013, the share of individual isolation increased from 20% to 23%. The study considers a person to be isolated when she has little or no social relations

 ISBN: 978-1-61208-325-4

with any of five following social networks: family, professional, friendship, and territorial affinities. Loneliness is expressed by a feeling of boredom, loss of meaning of life and uselessness. If loneliness can affect people with healthy relationships (4.1%), it is felt most when the person is alone (11.5%).

Health experts point out that loneliness has consequences on the health of the person since it creates depression, hygiene problems, various somatic diseases.

In response to this problem, public and private sectors focus on two types of intervention. The public sector focuses on the implementation of an organization that relies on communities of professionals or volunteers in a pre-determinated, organized manners. The conclusions of Monalisa report [7] to Michelle Delaunay, Minister Delegate to the Minister of Social Affairs and Health, in charge of the elderly and autonomy in France, reflect this approach. It recommends a regular schedule of visits to old people by teams of local volunteers. The private sector develops technology solutions and telecommunication security, installed in the person's home. These are designed from the geolocation techniques and objects identification and telecommunications. These facilitate the transmission of messages and detect unusual movements (falls, etc.). Robotics, in the form of androids, are now being added to these technologies; they look similar to humans in order to facilitate human interaction. They reproduce the behavior and expressions, vocal inflections, and even are be able to move hands, etc.

Both of these approaches are limited by their inability to take into account the complexity of the relationship, even if they can generate a network of solidarity and caregivers for the person alone or enhance interactions and alertness.

Monalisa Report [7] states: "To overcome the lack of family relationships, friends or neighbors, people with loss of autonomy and who feel lonely, try to forge friendships with caregivers and housekeepers they did not choose and who must adopt professionals' postures. This emotional dependence is rarely satisfactory and difficult to live on both sides".

Sometimes, the accumulation of interventions triggers the feeling of loneliness. Catherine Audibert [8] criticizes the services organization characterized by: "The fragmentation of cares" that "... not only does not permit therapeutic projection to a person of choice but keeps the patient in distress of loneliness."

The organizational and technological can only formulate programmed models. They are limited they cannot adapt on a variety of cultural and social contexts, life stories and ruptures biographical, emotions and motivations of personal life.

It appears that these models could be integrated into a more comprehensive approach, which reinstates the centrality of the individual. It seems necessary to ask about the problem of relational dynamics itself. From social ties to the emergence of friendly cooperation, it is question to build a relationship with expectations, needs, and values of the person.

IV. Of Friend Ships to Empathy

The study of the forms of friendship not only reveals a kaleidoscope of elements but also and especially its vital necessity for humans. Claire Bidart [9] who studied the structuring contexts of friendship emphasizes the important role of identity projection in friendship: "The network of friends would reflect the multiple identities of being, in which relations can play their ambivalences, their identities ambiguities."

Friendship appears as a space of self-reflexion that allows us to exist in the eyes of others and gives us the opportunity to fully experience our intimate part. Catherine Audibert [8] highlights how emotional and elective relationship with a chosen person allows a "serene solitude" that is to say, freedom to "grow its own solitary garden."

Friendship raises the relationship to oneself, through the other, as a participant in self-estime. The self-awareness with that part of intimacy, never revealed, is developed through the other one. It allows us to experience ourselves as existing in our uniqueness.

Finally, Serge Tisseron [10] emphasizes the comforting aspect of friendship. He notes: "The more we move away from what we look like to go to what is different from us, the more psychological stability is testing."

Friendship generates awareness of each other and by extension the sensation of an external world, with which one exist because, as Merleau Ponty [11] writes: "The world is not a subject, which I have the law of the constitution, it is the natural place and scope of all my thoughts and all my explicit perceptions (...) the man lives in the world, it is in the world he knows himself. ".

Perceive the other is to understand and feel him. The relationship with the other enhances the emotional, affective and cognitive abilities. Serge Tisseron [10] says this empathetic knowledge is composed of four dimensions: "The ability to feel the emotions of others, the ability to have a mental representation of its thoughts, the tendency to imagine oneself as different characters reals or fictionnals, concern for the other, which mobilizes support behaviors."

From these initial observations, the approach would be to establish the basis for a reflection on the possibilities of reconstruction of relational models around the elderly.

Older people are not without of these desires to communicate, but they prefer "elective" relations [8] in the sense that they are centered primarily on common affinities. About the needs of the elderly, Genevieve Arfeux-Vaucher, Director of Research at the National Foundation for Gerontology Paris, notes that: "It is more appropriate to speak of a social role rather than social ties".

V. Contribution of Social Networks to Socialization

We note the singular coincidence between the wishes of the elderly and those of websocial users. For Bernard Spiegler [12], the success of social networks (Facebook, Google +, web 2.0, sharing sites, etc.), shows relationships lacking in a highly individualized and fragmented society. Social networks are a necessary socializing invention. By

Leroi-Gourhan, Bernard Spiegler considers that the technology is anticipating needs.

McLuhan [13] wrote: "Any extension of human faculties is the reaction to irritation caused by the environment and comes in the form of requirements (...) The new medium is a drug for the save in the social balance. »

Bernard Spiegler [12] highlights how "relational technologies' answer inherent need for everybody to participate in the class. Philosopher Gilbert Simondon [14] considers that each individual part is connected to a "collective individuation," because "The unity of life is the whole group and not the isolated individual."

Technologies are often perceveid as a factor of dissolution of social ties. This common perception is included in the Monalisa [7], which reports: "these precarious relationships are a (...) phenomenon associated with the rise of individualized behavior, (...), the role played by television Internet and at the expense of direct social relations ... ".

Instead, Web 2.0 becomes vector relationships and interactions empathic level. Jeremy Rifkin [15] writes that through the mass use of social networks "... we learn that human nature is not to seek independence (...) but rather company, affection, and privacy. » He added: "The empathic approach is the existential awareness of the vulnerability we all share."

VI. REDEFINING FORMS OF SOCIALIZATION

We seek to redefine forms of resocialization of the elderly according to the individual requirement, personal relationships, taking into account their physical disabilities (loss of hearing, sight, the mobility, etc.) and their levels of vulnerability.

I would be possible to consider a complementary approach based on formalization of social networks.

It means to recognize the limits of the distance introduced by virtual communication but also to know its potential. Pierre Boutinet [16], with reference to Lucien Sfez [17] asks: "With communication, contrary to what happens in interpersonal relationships, finally the most important is the technical medium, which separate as it unites. This medium has a connected function tends to build screen playing fascination mirror game that transmitter and receiver each to his own loneliness. "

This function makes necessary an analytical approach to direct and indirect relational modes in order to consider the degrees of relational opportunities depending on the chosen media and contexts observed.

Bernard Stiegler [12] notes that social networks do not rely on deep friendships but increase the "relationships" or "vectors reputation", multiple "weaks links" made public in a "declarative". They obey specific relational rules. In this context, the need for elective affinities should be reconsidered.

VII. CONCLUSION

The objective of the paper is to consider scenarios of communication and interaction through social spaces defined by Catherine Delory-Momberger [18] "as spaces of growing together", which could be virtual and face to face. In the first case, the use of relational technologies may be required but appropriately integrating a set of elements related to lifestyle. In sum, our hypothesis is not so much to consider about the elderly a network of solidarity but to convince them to develop their own emotional network and reconstruct a confortable community, around their. Collecting data with narrative methods and analyzing it with tables based on different criteria can contribute to thinking of another way to consider the treatment of loneliness.

REFERENCES

[1] A. Muchielli, "Situation and Communication", Ed. Ovadia, Nice, 2010, p.170.

[2] F.Bernard, "ICT, discipline openness and decompartmentalization", in A. Bouzon: "Organizational communication debate. Fields, concepts, perspectives", Paris, L'Harmattan, 2006, pp. 33 - 46.

[3] J.L. Le Moigne, "Constructivist Epistemologies", Ed. PUF, 1995

[4] J. Jouet, "Critical review of the sociology of uses, Networks", n° 100 Paris, CENT / Hermès, 2000, pp. 487 – 521.

[5] N. Heinich, "Committed to neutrality", Communication issues, Records 3, 2006, pp. 83 – 96.

[6] Study "Loneliness in France", Fondation de France, Paris, June.2013.http://www.fondationdefrance.org/Outils/Mediatheque/Etudes-de-l-Observatoire

[7] J.F. Serres, "Recommendation for a national mobilization against social isolation of older people", Report Monalisa, Paris, July 2013.

[8] C. Audibert , "The inability to be alone, essay on love, loneliness and addiction", Ed. Payot, Paris, 2008, pp. 98 - 139.

[9] C. Bidart, "Friendship, social ties", Ed. La découverte, Paris, 1997.

[10] S.Tisseron, "Empathy in the heart of the social game", Ed. Albin Michel, Paris, 2010, pp. 75.

[11] M. Merleau Ponty, "Phenomenology of perception", Ed. Gallimard, Paris, 1945.

[12] B.Stiegler, "The most valuable at the time of sociotechnologies", Social networks, Ed. FYP, 2012, pp.14-36.

[13] M. Mac Luhan, "From eye to ear, the New galaxy" Ed. Hartubise, Montréal, 1977.

[14] G.Simondon, "The individual and his hysico-biological genesis", Ed. Jérôme Milliam, Paris, 1997, pp. 156.

[15] J. Rifkin, "A new conscience for a world in crisis, to a civilization of empathy" , Ed. Actes Sud, 2011, pp. 33.

[16] J.P. Boutinet, "The immaturity of adulhood", PUF, Paris, 1998.

[17] L. Sfez , "Critical communication ", Ed. Seuil, Paris, 1988.

[18] C. Delory – Momberger, "The subject in the city, social policies in the territories and spaces recognition", Ed. Plein Feux, Paris, 2009.

Transparent Electrostatic Actuator with Mesh-structured Electrodes for Driving Tangible Icon in Tabletop Interface

Ryuta Iguchi, Takuya Hosobata and Akio Yamamoto
Department of Precision Engineering, School of Engineering
The University of Tokyo
Tokyo, Japan.
{r.iguchi, hosobata, akio}@aml.t.u-tokyo.ac.jp

***Abstract*—This paper proposes a transparent electrostatic actuator for actuating tangible objects in a tabletop interface. The actuator is a planar actuator with two degrees-of-freedom, capable of actuating a small dielectric sheet placed on top. The feature of the actuator is in having mesh-structured electrodes, patterned on a transparent film using a conductive polymer. By applying pulse voltages to the electrodes, electrostatic force act on the charges induced on the sheet to produce step-wise motion. The transparency of the actuator enables this actuation while maintaining the visibility of graphics on a visual display underneath. To demonstrate the suitability of the actuator for a tabletop interface, we constructed a simple human-computer interaction system, in which a physical icon travels on the actuator placed on a flat panel display. In the demonstration, the user could interact with the computer by handling active and passive tangible icons. The actuator used in the system could actuate a small icon in a broad area of 300 mm by 300 mm, with application of 500 V pulse voltages.**

Keywords-tangible media; tabletop interface; human-computer interaction; surface actuator; electrostatic actuator.

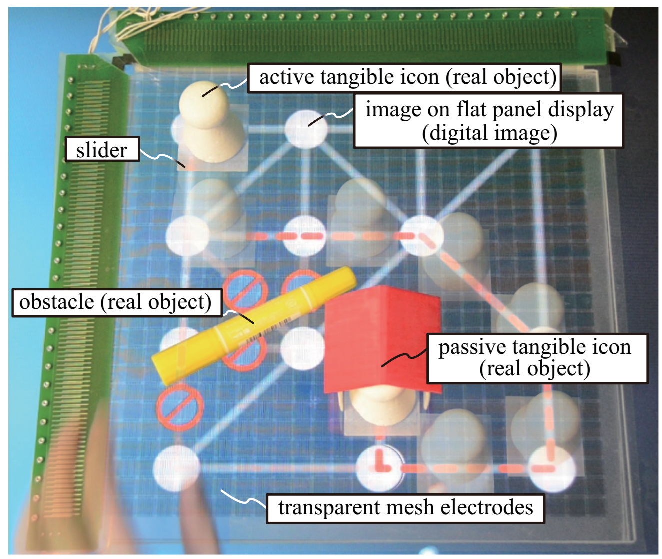


Figure 1: Conceptual photograph of tabletop interface on flat panel display with tangible icon actuated by mesh-structured electrostatic actuator.

I. Introduction

In 1997, Ullmer and Ishii showed that great enrichment is brought to a tabletop interface by bringing in physical objects as tangible media to coexist with a visual display [1]. A way to further improve such an interface was demonstrated by Pangaro et al., which is to have the tangible objects actuated [2]. Actuated tangibles let the computer inform us of its internal state through their physical motion, which we can comprehend with familiarity. In this paper, we introduce an actuator developed for such tabletop interface: a transparent electrostatic actuator with mesh-structured electrodes. In addition, we prove its effectiveness through a simple tabletop interaction, which is illustrated in Fig. 1. The actuator specializes in actuating physical icons on a surface of a Flat Panel Display (FPD).

Forms and functions of active tangibles vary depending on types of interactions expected in the tabletop interface. To satisfy the requirements of each case, it is important to choose the proper means of actuation. To have the active tangibles represent handheld-sized objects, the most effective way is to have the objects move by themselves, that is, to use wheeled robots [3][4] or legged robots [5]. An advantage of this solution is that multiple robots can collaborate to perform various tasks. However, shapes and sizes of the objects representable with robots are limited by their functionality; the robots must carry batteries and driver electronics, which prevents the robots to disguise as physical icons that do not resemble them.

To have physical icons of various shapes actuated, a tabletop interface must incorporate an actuator capable of driving objects on a displayed image. Various magnetic surface actuators satisfy this requirement, which include electromagnet arrays [2][6], planar induction motors [7] and sawyer motors [8]. Magnetic levitators are also effective, providing an actuation in mid-air [9][10][11]. These actuators have excellent driving performance, but the drawback is in requiring ferromagnetic cores embedded in the system, which consume space and increase the total weight of the system.

As another choice of actuator for actuating physical icons, we propose a transparent electrostatic actuator, which we can easily install on an FPD. The original concept of the transparent actuator was proposed by Egawa et al. in 1991 [12]. The actuator is a transparent plastic film with array of strip-electrodes made of transparent conductors. The film drives dielectric sheets (e.g., papers, plastics, and glass) placed on its surface, by electrostatic forces excited by voltages applied

 ISBN: 978-1-61208-325-4

to the electrodes. The actuator is suited for tabletop interface because of its transparency and thinness; simply by placing it on top of a visual display, we can drive objects while maintaining visibility of displayed images.

Previously, we reported an FPD-based interaction using the transparent electrostatic actuator, in which users could play catch with an animated cat, using a ball printed on a paper sheet [13]. In this system, the actuator could actuate the ball only in one direction. To realize multi-directional motion, we developed a transparent actuator with 2-DOF [14], and demonstrated a mixed-reality on an FPD, in which an animated robot could interact with a small ball driven based on virtual dynamics [15]. One problem of this actuator was in having short stroke; it could move the ball only within a small area of the FPD, because the actuated film was much larger with respect to the ball it carried. Such restriction in stroke was caused by the actuator's electrode structure. Thus, as a next step, we chose to develop a 2-DOF transparent actuator that can actuate small films in a broad area.

The transparent electrostatic actuator we developed in this work has mesh-structured electrodes, which realize actuation of small objects with large stroke. The idea of using mesh as electrodes in an electrostatic actuator is already reported using woven copper wires [16]. We combined this idea with the concept of transparent actuator, and developed an actuator with *printed* mesh, fabricated by screenprinting using a transparent conductive polymer. In the mesh, there are vertical and horizontal *printed* belts, containing three-phase electrodes. The belts intersect in a manner that electrodes producing different directioned forces appear on the surface alternately. When the electrodes are excited with pulse voltages, electrostatic forces arise to give a step-wise motion to a dielectric sheet placed on top. We designed the belts to be 12 mm in width, which makes the minimum size of the driven film twice this size. The actuator itself has an area of 300 mm by 300 mm, which allows long distance travel of the actuated sheet. With the long stroke and the transparency, the actuator enables planar actuation on visual displays, which is especially useful for tabletop interfaces.

The rest of the paper is organized as follows. In Section II, we explain the structure and driving principle of the transparent actuator. Then, we report on the evaluation of the actuator's performance in Section III. Finally, in Section IV, we demonstrate the actuator's suitability for tabletop interfaces through a simple human-computer interaction with a physical icon driven on an FPD.

II. Transparent Electrostatic Actuator with Mesh Electrodes

This section provides explanations on the structure and the driving principle of the transparent mesh actuator shown in Fig. 2. Details on the principle of 1-DOF actuator is reported in [12]; the same principle applies to this actuator, but with an extra DOF.

A. Structure of the Actuator

Fig. 3 shows the basic structure of the actuator. The actuator consists of two films; we refer to the fixed film as the stator and the driven film as the slider. The stator has a polyethylene terephthalate (PET) film with 100 μm

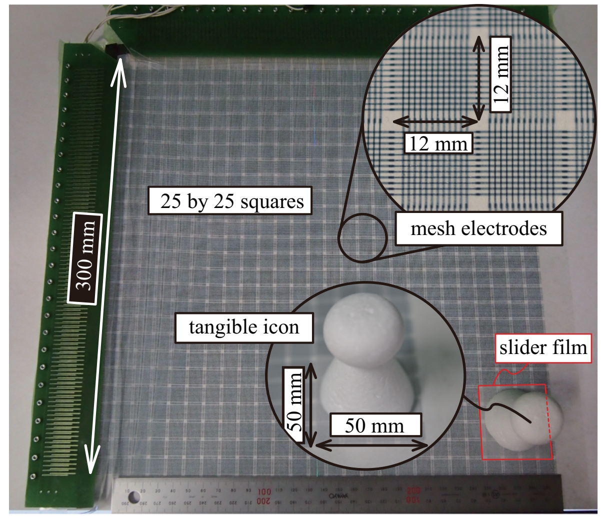


Figure 2: Photograph of the transparent electrostatic actuator with mesh electrodes.

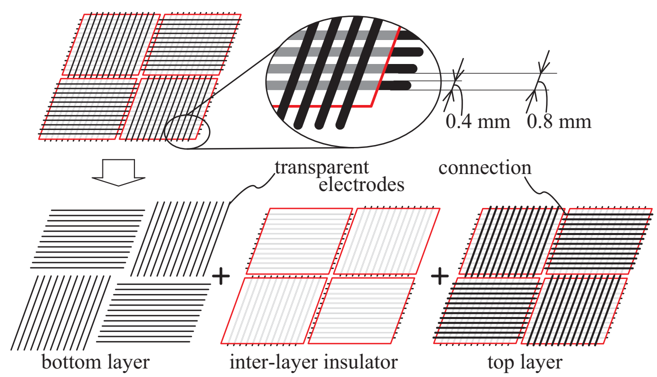


Figure 3: Structure of the transparent electrostatic actuator with mesh electrodes.

thickness as a substrate. On the surface of this film, patterns of electrodes and insulators are printed by screenprinting technology. The electrodes are strips of transparent conductive polymer: poly(3,4-ethylenedioxythiophene):poly(4-styrenesulfonate) (PEDOT:PSS), 400 μm in widths and spaces, printed in two layers. We chose this polymer because of its transparency and compatibility with the screenprinting; the polymer appears in pale blue, but is transparent enough to see through the actuator.

The electrodes in the bottom layer are directly printed on the PET film. There are vertical electrodes and horizontal electrodes; the former are the X electrodes and the latter are the Y electrodes, which respectively drive the slider in directions X and Y. The X and Y electrodes are separated in small squares, arranged alternately as in a checkerboard pattern. Note that in this layer, electrodes in all squares are isolated from each other. Lengths of the sides of each square are 12 mm, and each contains twelve electrodes. There are 25 by 25 squares in total, which makes the total length of the stator 300 mm.

On top of the bottom-layer electrodes, there is a inter-layer insulator, which is also screenprinted using thermo-setting polymer paste. The insulator covers most of the bottom-layer electrodes, but leaves the ends of the electrodes uncovered.

Over this inter-layer insulator, top-layer electrodes are printed. These electrodes are arranged in different directions from the bottom-layer electrodes; X electrodes are printed over Y electrodes and vice versa. The top-layer electrodes are connected at their ends to the bottom-layer electrodes, in areas not covered by the inter-layer insulator. Thus, the connected X and Y electrodes form 300 mm long belts, which appear as if they are woven, with X and Y belts alternately appearing to the surface. Finally, the surface is covered with another insulating layer, on which the slider is driven.

The stator is very thin; each of the insulating layers are approximately 20 μm thick, which makes the total thickness of the stator 147 μm. Thickness of the electrodes are 2 μm, almost negligible compared to the film thickness. The thinness, together with the transparency, provides a feeling that the slider is driven directly on the images displayed underneath.

As a slider, we could choose from various dielectric sheets, such as paper, plastic or glass. In this work, we chose to use a slider with PET sheet 50 mm $\times$ 50 mm as a substrate, with its top surface coated with glass particles 100 μm in diameter. This slider was chosen empirically; the slider performed the best among those we tested. There is an interesting problem in the optimization of materials for the slider, which we leave for future works.

B. Driving Principle

Fig. 4 shows the patterns of voltages applied to the electrodes to drive the slider in different directions. The X and Y electrodes in the outer most squares are connected to three-phase buses, to which three-phase pulse voltages are applied. The slider is driven in the X direction by shifting the pulse voltages on X electrodes while the electrodes of Y are grounded, and the similar applies to driving in the Y direction. To drive the slider in a diagonal direction, we shift pulses on both X and Y electrodes.

Fig. 5 schematically illustrates the induced charges on the slider and the electrostatic forces acting on them. The driving sequence consists of three steps: charging, shifting and stepping. First in charging, the electrostatic field excited by the applied voltages induce charges of corresponding polarity on the slider. Then the voltage pattern is shifted in the driving direction, but the arrangement of charges lags behind because the slider has a low conductivity. Finally, electrostatic forces act on the lagged charges and steps the slider forward.

For the slider used in this work, 10 to 30 seconds of pre-charging was necessary as a preparation for driving. This is because the time constant for the charges to arise or decay was on the order of seconds. Once the slider was pre-charged, constant shifting of voltages could drive the slider full-stroke in one direction. However, occasional re-charging was necessary when the slider had to turn 90 degrees. For example, without this re-charging, the slider could not move in Y direction after traveling full-stroke in X direction. The reason is still unclear, but we presume that the charges required for driving in Y direction diminish while the slider is traveling in X direction.

The stepping distance of the slider depends on the balance

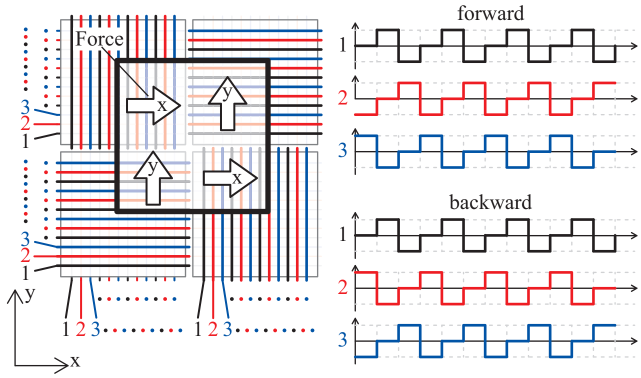


Figure 4: Voltage patterns applied to the electrodes in the mesh actuator.

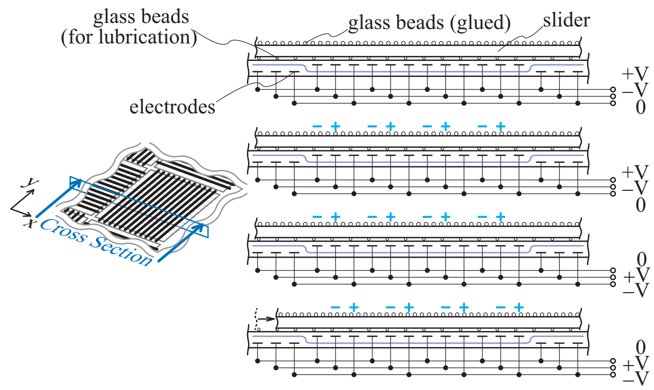


Figure 5: Schematic of driving principle of the mesh actuator.

of thrust and frictional force. With the actuator of this work, the friction is small enough and the stepping distance per pulse is nearly equal to the length of an electrode pitch: 800 μm. This characteristic allows us to roughly control the position of the slider in open-loop. In the application in Section IV, we take advantage of the characteristic to realize variety of motions with a simple control method.

III. Performance Evaluation

We conducted two experiments to evaluate the actuator's performance. First, we tested for the maximum speed the slider could travel. Then, we confirmed the actuator's capability of actuating the slider over all areas of the stator in different directions.

sends it to a DSP (DS1104, dSpace [17]) in Fig. 2, which is made of a polystyrene foam and weighs 0.48 grams. The magnitude of the pulse voltage was fixed at $\pm$500 V and glass beads of 20 μm diameter were scattered on the surface of the stator to reduce friction. Before each experiment, the slider was pre-charged by applying $(500, \ -500, \ 0)$ V DC voltages to both of X and Y electrodes for 30 seconds.

A. Maximum Speed

We drove the slider in X direction with pulse voltages of different frequencies, to see how fast the slider can travel while carrying the icon. Fig. 6 shows the motion of the slider driven

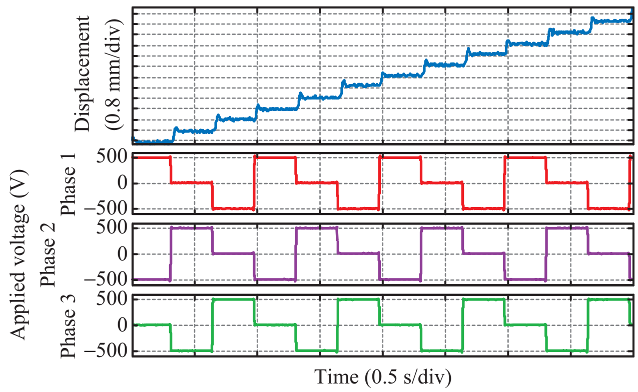


Figure 6: Step-wise motion of slider carrying icon and applied three-phase pulse voltage.

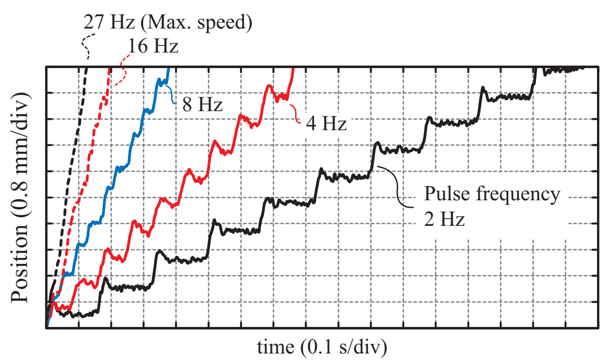


Figure 7: Motion of the slider driven by 500 V three-phase pulse voltages with different frequencies.

by 1 Hz pulse voltage in X direction, recorded using laser displacement sensor (Keyence LB-01). This result shows that the slider travels 2.4 mm per one cycle of the three-phase pulse, which is equal to the length of three electrode pitches. Fig. 7 shows the results with higher frequencies. The results indicate that the step-wise motion is damped as the speed increases with the frequency. This is because the inertia of the slider and the friction disable the slider to instantly respond to the pulsating force.

When the frequency exceeded 27 Hz, the slider stood still at its initial position; the maximum speed was approximately 64 mm/s. The speed is nearly equal to the frequency times three electrode pitches ($27\ \text{Hz} \times 2.4\ \text{mm} = 64.8\ \text{mm/s}$). This indicates that even at the maximum speed, the traveling distance of the slider is controllable by the number of voltage pulses. Thus, we can assume that the position of the slider is roughly controllable in open-loop.

B. Full-stroke motion

To confirm the actuator's capability as a long-stroke surface actuator, we recorded the full-stroke motion of the slider in different directions. In this experiment, the frequency of the pulse voltage was fixed at 5 Hz, and trajectories of the slider were recorded using a visual motion tracking system with high-speed camera (VW-6000, Keyence). The results are shown in Fig. 8, indicating that slider was successfully driven over the

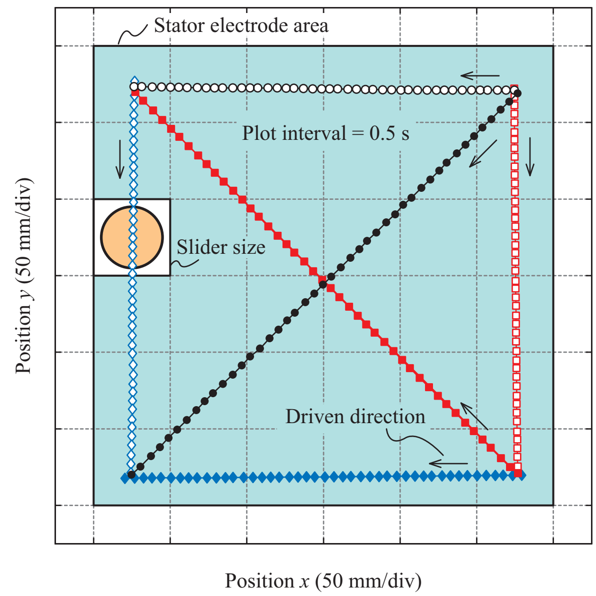


Figure 8: Result of Motion Tracking.

whole 300 mm × 300 mm area of the stator.

We expect that driving in intermediate angles (such as 30 degrees) is possible in principle, by exciting X and Y electrodes with pulse voltages of different frequencies. However, we could not realize such motion in this work; cause of this is still under investigation.

IV. Tangible Interaction Using Mesh Actuator

To demonstrate the actuator's suitability for tabletop interfaces, we constructed a simple interaction system illustrated in Fig. 9. In this system, the transparent mesh actuator covers an FPD, which is displaying an undirected graph. The actuator drives the person icon carried by the slider along the shortest path leading to another icon, which represents the person's home.

The user of this system can take three actions to interact with the computer. First, the user can pick up and move the person icon. Second, the user can move the home icon to change the destination of the path. In addition, the user can block the path by placing an obstacle (a marker pen). All these actions are monitored by the computer, as it is constantly tracking the positions of the objects using a web camera(PlayStation Eye, Sony Computer Entertainment). The computer responds to the actions by rerouting the path, and the result is presented to the user not only by the displayed images, but also by the person icon's actual motion.

A. Control Approach

Fig. 10 shows how the person icon is controlled during the interaction. The person icon and the home icon have their own nodes aside from those in the preset graph, and these nodes are automatically connected to nearby links.

The computer constantly updates the shortest path from the person icon to the home icon using the Dijkstra's al-

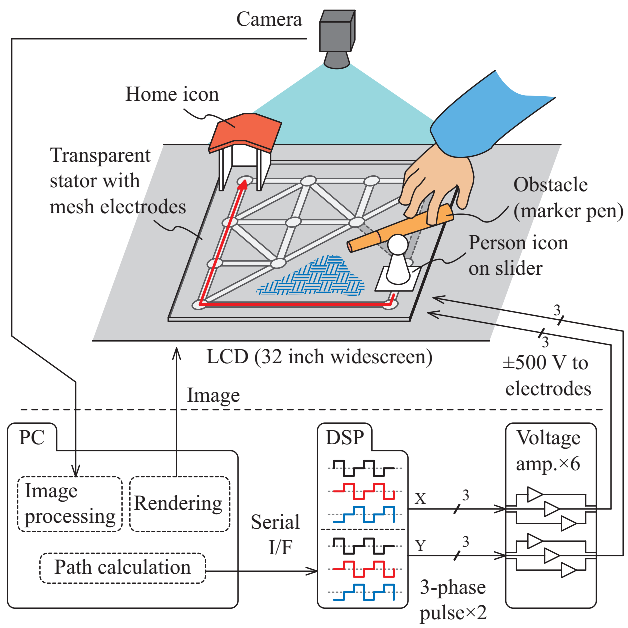


Figure 9: Configuration of the tangible tabletop interface using transparent mesh electrostatic actuator.

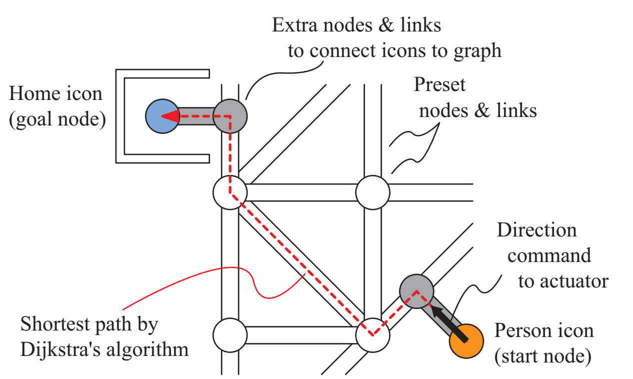


Figure 10: Control approach to drive the slider along the shortest path.

gorithm. Based on the calculated path, the computer detects the direction from the person icon to the next node in the path, and sends it to a DSP (DS1104, dSpace [17]). The DSP generates waveforms of three-phase pulses corresponding to the direction. The pulses are amplified by High-Voltage Amplifiers (HVA4321, NF Co. [18]) and applied to the electrodes to drive the person icon in the specified direction.

An exceptional condition is when the person icon is too close to the next node. In such a condition, the computer uses the direction from the person icon to the node after the next. This exceptional procedure was necessary, because without it the person icon would stop advancing and start oscillating around the next node. In this work, we applied the procedure when the icon was within a 10 mm diameter circle of the next node.

We chose this simple control approach because the mesh actuator can generate step-wise motion. This motion ensures that the driven icon does not overshoot beyond the next node by long distance. With this characteristic, specifying direction was sufficient to keep the driven icon on the planned path. Such simplicity in control is one advantage of this actuator, especially for tabletop interfaces using visual tracking; realizing fast servo control is difficult with cameras, whose frame rate limits the speed of the control cycle.

B. Motion of the Slider During Interaction

Fig. 11 shows the motion of the person icon recorded by tracking camera during three interaction experiments. In these experiments, the computer responded to the three actions from the user by rerouting the driven path of the person icon. In each experiment, the slider film carrying the person icon was pre-charged for 30 seconds by applying $(500, -500, 0)$ V DC voltages to both X and Y electrodes, and then driven at 7.2 mm/s by a ± 500 V three-phase pulse voltage. As explained in Section II-B, the slider film was re-charged for 10 seconds every time it had to make a 90 degrees turn. The re-charging was done by applying the same set of voltages as in pre-charging.

Fig. 11a shows the result of the user repositioning the person icon. At first, the person icon was traveling from the upper-right corner toward the home icon, along the shortest path in the graph. Then, the user picked the person icon up and repositioned it to prompt the computer to reroute the path, along which the person icon continued its travel. Fig. 11b and Fig. 11c show the result of the user's indirect action of handling passive tangible objects. In Fig. 11b, the user extended the travel of the person icon by changing the position of the home icon, which is the destination of the path. Fig. 11c shows the result of placing a marker pen: the obstacle, to block the preplanned path. The computer immediately detected that the links in the preplanned path is no longer available and rerouted the path.

C. Problems

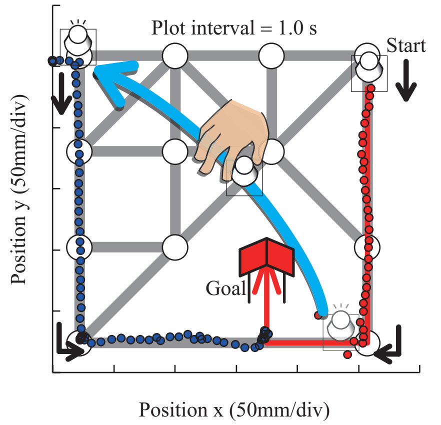
The results prove that the proposed actuator realizes a variety of human-computer interaction by allowing active (driven) and passive tangible objects to collaborate with computer graphics on a surface of an FPD. However, through these experiments, we also found some problems of the actuator.

First, the charging time needs to be shortened, because the waiting time during the charging prevents the user from being immersed in the interaction. Second, the gripping force of the slider need to be increased. The user could touch and feel the person icon, but was provided with no force feedback, because of the low friction, light weight and weak force. These problems indicate that further improvements are necessary in terms of materials, design and driving method of the actuator.

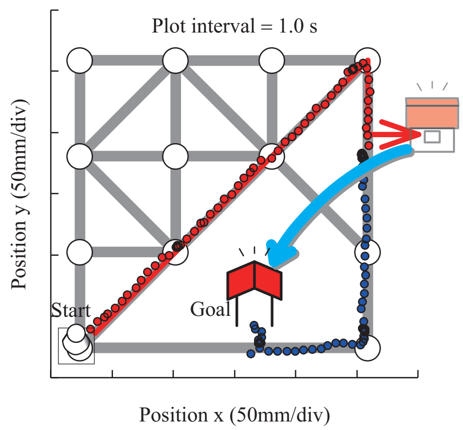
V. Conclusion and Future Works

In this paper, we proposed a transparent electrostatic actuator with mesh-structured electrodes especially designed for tabletop interaction. With a working prototype, we proved the actuator's capability of presenting various planar motions of a physical icon together with graphics displayed on an FPD underneath the actuator. To extend the actuator's application

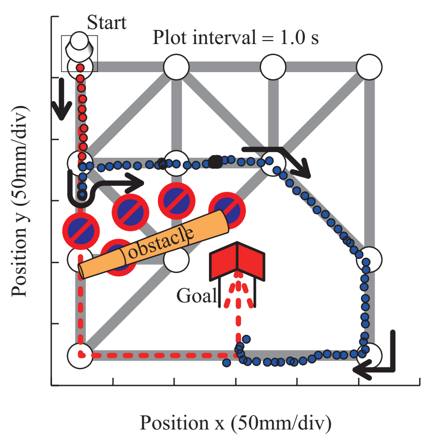
 ISBN: 978-1-61208-325-4



(a) Repositioning active icon (person icon).



(b) Changing destination by moving passive icon (home icon).



(c) Path blocking by placing obstacle (marker pen).

Figure 11: Rerouted path of person icon by user's handling of tangible icons. Plots indicate trajectory of person icon recorded with tracking camera.

range, a challenge lies in simultaneous actuation of multiple objects. In future works, we expect to realize this by independently controlling the voltages on different sections of the actuator.

ACKNOWLEDGMENT

This work was supported by Funding Program for Next Generation World-Leading Researchers (#LR013) from JSPS, Japan.

REFERENCES

[1] B. Ullmer and H. Ishii, "The metaDESK: models and prototypes for tangible user interfaces," In Proc. of the 10th Annual ACM Symposium on User Interface Software and Technology, 1997, pp. 223–232.

[2] G. Pangaro, D. Maynes-Aminzade, and H. Ishii, "The actuated workbench: computer-controlled actuation in tabletop tangible interfaces," In Proc. of the 15th Annual ACM Symposium on User Interface Software and Technology, 2002, pp. 181–190.

[3] E. W. Pedersen and K. Hornbæk, "Tangible bots: interaction with active tangibles in tabletop interfaces," In Proc. of the 2011 Annual Conf. on Human Factors in Computing Systems, 2011, pp. 2975–2984.

[4] M. Sugimoto, T. Fujita, H. .Mi, and A. Krzywinski, "RoboTable2: a novel programming environment using physical robots on a tabletop platform," In Proc. of Int. Conf. on Advances in Computer Entertainment Technology, 2011, pp. 10:1–10:8.

[5] S. Somanath, E. Sharlin, and M. C. Sousa, "Integrating a robot in a tabletop reservoir engineering application," In Proc. of ACM/IEEE Int. Conf. on Human-robot Interaction, 2013, pp. 229–230.

[6] M. Weiss, F. Schwarz, S. Jakubowski, and J. Borchers, "Madgets: actuating widgets on interactive tabletops," In Proc. of the 23nd Annual ACM Symposium on User Interface Software and Technology, 2010, pp. 293–302.

[7] S. Yoshida, H. Noma, and K. Hosaka, "Proactive Desk II: Development of a New Multi-object Haptic Display Using a Linear Induction Motor," In IEEE Virtual Reality Conf., 2006, pp. 269–272.

[8] M. Sugi et al., "Motion Control of Self-Moving Trays for Human Supporting Production Cell "Attentive Workbench"," In Proc. of the IEEE International Conf. on Robotics and Automation, 2005, pp. 4080–4085.

[9] J. Lee, R. Post, and H. Ishii, "ZeroN: mid-air tangible interaction enabled by computer controlled magnetic levitation," In Proc. of ACM Symp. on User Interface Software and Technology, 2011, pp. 327–336.

[10] M. Marshall, T. Carter, J. Alexander, and S. Subramanian, "Ultra-tangibles: creating movable tangible objects on interactive tables," In: Proc. of SIGCHI Conf. on Human Factors in Computing Systems, 2012, pp. 2185–2188.

[11] P. Berkelman, M. Miyasaka, and J. Anderson, "Co-located 3D graphic and haptic display using electromagnetic levitation," In: Proc. of Haptics Symp., 2012, pp. 77–81.

[12] S. Egawa, T. Niino, and T. Higuchi, "Film actuators: Planar, electrostatic surface-drive actuators," In Proc. of the IEEE Micro Electro Mechanical Systems Workshop, 1991, pp. 9–14.

[13] K. Amano and A. Yamamoto, "Tangible interactions on a flat panel display using actuated paper sheets," In Proc. of the ACM Int. Conf. on Interactive Tabletops and Surfaces, 2012, pp. 351–354.

[14] T. Hosobata, A. Yamamoto, and T. Higuchi, "2-DOF Synchronous Electrostatic Actuator with Transparent Electrodes Arranged in Checkerboard Patterns," In Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, 2013, pp. 4919-4924.

[15] T. Hosobata and A. Yamamoto, "Mixed Reality System on Flat Panel Display with Real Object Driven by Synchronous Transparent Electrostatic Actuator," In Proc. of Int. Conf. on Multimedia and Human Computer Interaction, 2013, pp. 127:1–127:7.

[16] A. Yamamoto, H. Yoshioka, and T. Higuchi "A 2-DOF Electrostatic Sheet Conveyer Using Wire Mesh for Desktop Automation," In Proc. of IEEE Int. Conf. on Robotics and Automation , 2006, pp. 2208–13.

[17] http://www.dspace.com/en/pub/home/products/hw/singbord/ds1104.cfm Last accessed on 10/Feb/2014.

[18] http://www.nfcorp.co.jp/english/pro/pp/p_amp/h_vol/hva4321/index.html Last accessed on 10/Feb/2014.

Interaction with Real Objects and Visual Images on a Flat Panel Display using Three-DOF Transparent Electrostatic Induction Actuators

Norio Yamashita, Kota Amano, and Akio Yamamoto
Department of Precision Engineering
University of Tokyo
Tokyo, Japan
Email: {norio, amano, akio}@aml.t.u-tokyo.ac.jp

***Abstract*—This paper describes a new type of computer-user interactions through an actuated real object and visual images on a liquid crystal display (LCD). The interactions are realized by three degrees-of-freedom (three-DOF) transparent electrostatic induction actuators placed on the display. The dynamic interaction was realized using asynchronous driving characteristics of the induction actuator. The actuator does not interfere with visual images on the screen because of the optical transparency of the actuator. The actuator is equipped with eight sets of one-DOF actuators, which enables three-DOF driving, XY translation and rotation, of sheet objects. Fundamental performances of the actuator regarding the 3-DOF motions are investigated. Also, a prototype interactive application is demonstrated.**

Keywords-electrostatic actuator; transparent; three-DOF; visual interaction;

I. Introduction

Recent progress of information technologies have provided us with many digital contents. In the use of the digital contents, user interfaces have been given an important role to determine user experiences. In the days when pointing devices and keyboards were principal input devices, touch panel screens enhanced user experiences by giving intuitive interaction with computers. While their use of touch modality is limited for manipulation of digital graphics, other trials of using real tangible objects for communication modality have also been studied [1–4]. In such interactions, users can communicate with computers intuitively by touching, moving, and picking up the real objects. This intuitive interaction will be further enhanced by returning responses to the real objects by actuating them based on the digital data, which requires embedded actuators. Most of the actuators used for such interaction systems are electro-magnetic actuators. Although electromagnetic actuators have high output forces and high controllability, they typically have heavy and bulky structures.

Recently, an electrostatic actuator has been introduced in this field as an alternative actuator [5–7]. The electrostatic actuator is a thin and lightweight actuator, which can directly drive a sheet-type object placed on it [8]. In addition, it can be manufactured in transparent color by means of transparent conductive materials, such as indium tin oxide (ITO). Since it does not occlude the images on a LCD display, combinations of the actuators and LCD displays have successfully demonstrated interactions with a real tangible object and virtual images, with considerably simple system setups.

The electrostatic actuators that were incorporated in those systems can be divided into two different types: asynchronous and synchronous types. As discussed more in detail in the next section, due to their different characteristics the suitable applications are also different. Asynchronous type is suitable for dynamic user interaction, whereas synchronous type is suitable for rather static one. For synchronous type, two-DOF motions have been demonstrated for active tangible systems [7], [9]. On the other hand, for asynchronous type, only one-DOF actuator has been introduced for active tangible systems, although three-DOF motion was already investigated for this actuator, apart from active tangible systems [10].

In this paper, a transparent electrostatic asynchronous actuator to realize three-DOF motion is reported and applied for an active tangible system. The system utilizes pen-tablet digitizer to detect the motion of an actuated sheet. With the help of the digitizer and the actuator, the system realizes intuitive interaction with computer animation through a sheet-like tangible object.

Section II compares two different driving characteristics of the electrostatic actuators from the viewpoints of interactive systems. Section III introduces the basic driving principle of an electrostatic actuator. Section IV describes the structure and features of fabricated transparent 3-DOF electrostatic actuator. Section V evaluates the actuator experimentally. Section VI proposes active tangible interaction system using the actuator. Section VII concludes this paper.

II. Electrostatic Actuators for Interactive Systems

Electrostatic actuators have several types of driving methods, which provide different characteristics of slider behavior. According to the interactive systems, we can classify them into two different types based on their driving characteristics: synchronous and asynchronous. In the synchronous type, the displacement of the actuated object is determined by driving signal. In the case of the synchronous electrostatic motor [7], [9], the displacement or speed of the object is in proportional relationship with the phase or frequency of the driving ac signals. This facilitates positioning of the object in open-loop control without any external sensors. By changing phase or frequency, the system can easily control the position or the speed of the object. However, when subject to too large external forces (which means overloading), synchronous actuators typically ”step-out”. If the external load is too large such that the actuator cannot keep the designated object displacement

 ISBN: 978-1-61208-325-4

TABLE I: FEATURES OF TWO DIFFERENT ACTUATION PRINCIPLES

Driving motion	Synchronous	Asynchronous
Position control	Available in open-loop control	Requiring Sensor feedback
Behavior in overloading	step-out and causes vibration	smooth motion without step-out
Suitable application	Applications requiring relatively accurate positioning	Dynamic interaction between users and tangible objects

anymore, the object will jump to some other point, which is called step-out. When an actuator steps-out, the object shows vibrating behaviors because of the jumping, which can be felt unpleasant if it is used in an interactive system.

In an asynchronous type, the driving force is determined by driving signal. The displacement of the object has no direct relation with the driving signal. Thus, for displacement or speed control, sensor feedback is necessary. On the other hand, there is no "step-out" in asynchronous actuator since asynchronous operation does not define the displacement. If there is too much external force, the object will be simply pushed back without distinct vibrations.

The two different actuation principles features are summarized in Table I. Due to their different characteristics, they can be used in different interactive applications. A synchronous type would be more suitable for applications that require position control. For example, the application demonstrated in [7] fully exploited the feature of the synchronous actuator. It realizes synchronous motions of computer animation and a tangible object, without any external sensors.

An asynchronous type would be rather suitable for applications requiring dynamic interaction. Here, "dynamic interaction" means that a user continuously moves the object to interact with some computer graphics. In [5], [6], such interactions have been demonstrated where a user can play a game of catch with a computer animation. In this application, the asynchronous behavior was imperative, as the user pushes back a ball against actuator's operation. If a synchronous actuator was utilized, the user would have felt unpleasant vibrations.

In this work, we focus on asynchronous actuator and develop 3-DOF version to realize more dexterous interaction than those demonstrated in [5], [6].

III. BASIC PRINCIPLE OF ELECTROSTATIC ACTUATOR

The three-DOF actuator used in this paper combines several one-DOF actuators. Thus, the basic structure and principle of one-DOF actuator is explained first. Fig. 1 illustrates the basic structure. It has a sheet with embedded electrodes, which is called "stator", and a sheet object that is to be actuated. The slider electrodes are connected into three phases so that three-phase ac voltage can be applied. The actuated object is a dielectric material without any electrode inside. A paper sheet or a plastic sheet is available as the actuated object.

Applying ac voltages to the stator electrode generates traveling voltage wave on the stator, which induces traveling electric charge wave on the slider, as shown in Fig. 1. The electrostatic interaction between the two traveling waves provides horizontal actuating force. As the electrostatic interaction also provides vertical attraction force, the friction between the actuated object and the stator must be kept low enough.

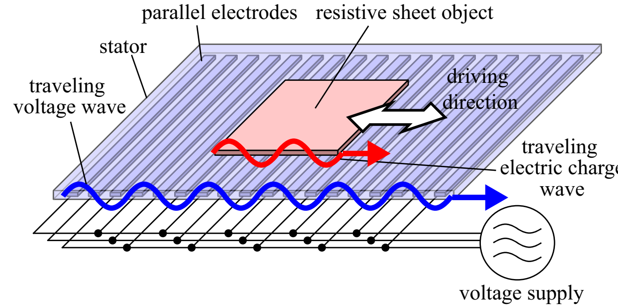


Figure 1: Schematic diagram of a one-DOF electrostatic actuator.

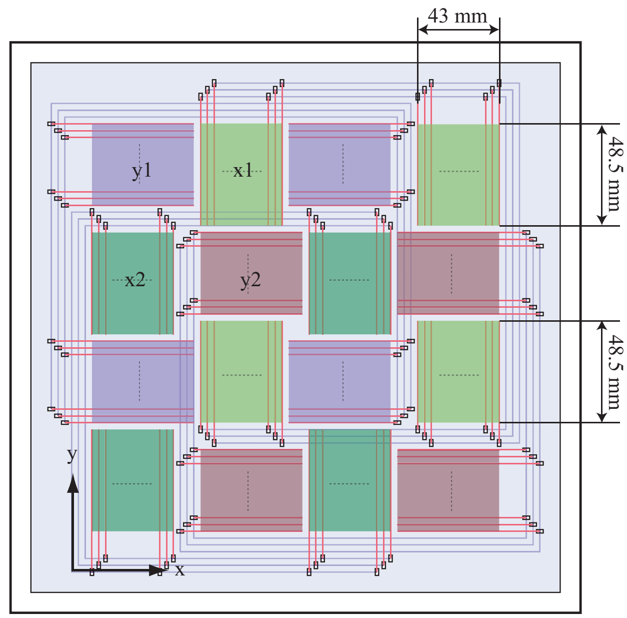


Figure 2: Structure of a 3-DOF actuator designed using combination of 1-DOF electrodes of X or Y.

IV. TRANSPARENT THREE-DOF ACTUATOR

A. Structure

This section describes a transparent three-DOF electrostatic actuator, which can realize XY translation motion and rotating motion of a slider. The three-DOF actuator is designed by tiling multiple one-DOF actuators in a checkerboard pattern. Fig. 2 shows the design of the electrodes, which has a four-by-four tiling structure. The units in the same color are connected to the same bus lines, and thus will be activated with the same voltage patterns. Fig. 3 shows driving methods for translations. One half of the electrodes produce X-direction force and the others produce Y-direction force. Using either one of the electrode groups, one can drive a slider in an X or Y translational motion. Diagonal motion can be available using both of them. The direction of the translation can be changed by exchanging any two of the three phases of the voltages, which reverses the traveling direction of the voltage wave.

Rotating motion can be produced using a combination of four electrodes, as shown in Fig. 4. Rotating can be realized only on limited points. As shown in a later section, possible rotating direction is fixed for each point due to the electrode structure. The rotation toward the other direction is not stable. Fig. 5 shows the available rotating points and their stable

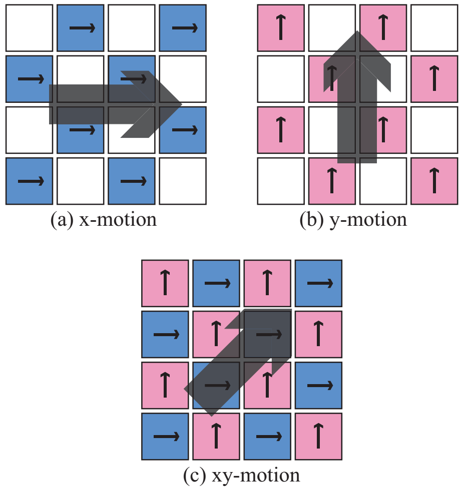


Figure 3: Driving method for X, Y, and diagonal translations.

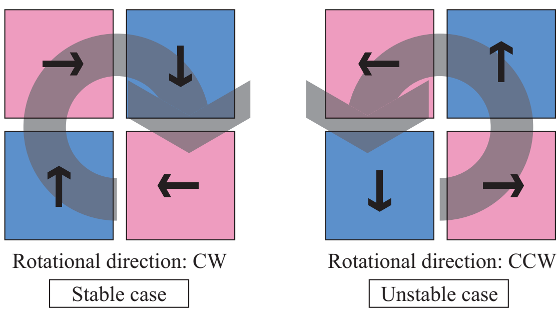


Figure 4: Driving method for rotation. (Left) stable and (Right) unstable rotation in a rotating point.

rotating directions. Four points have clockwise (CW) rotating direction and the others have a counter-clockwise (CCW) rotating direction.

This rotating motion is unique to this asynchronous actuator; it is not possible with the two-DOF synchronous actuator reported in [7] as the synchronous actuator requires the electrodes within an actuated object to always be parallel to the stator electrodes.

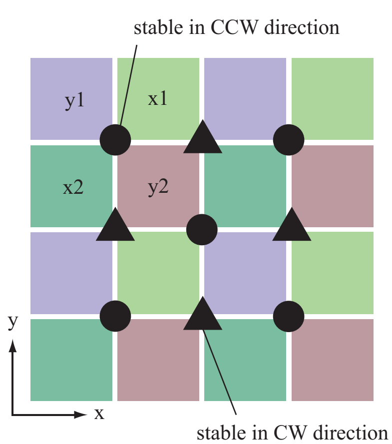


Figure 5: Available stable clockwise (CW) or counter-clockwise (CCW) rotations in a 3-DOF electrode.

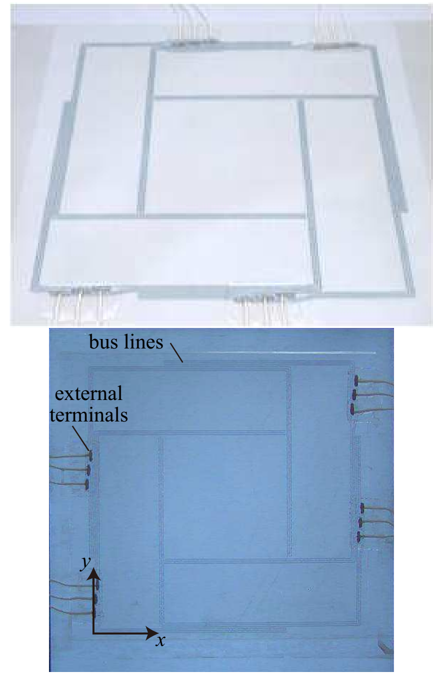


Figure 6: Photo of a 3-DOF transparent electrostatic actuator (upper) and its appearance on an LCD display (lower).

B. *Fabrication*

Fig. 6 shows the transparent 3-DOF actuator that was developed in this work. The actuator has transparent ITO electrodes on polyethylene terephthalate (PET) sheet, which is also transparent [6]. Although the main actuating electrodes are transparent ITO, bus lines, which connect the actuating electrodes into three phases, are not perfectly transparent as it is made of conductive polymer ink; it has pale blue color.

In the fabrication of the actuator, three-phase actuating electrodes were fabricated on a PET sheet by etching. Then, they were covered with insulating ink except for through-hole positions by screen-printing. Next, bus lines were printed over the through-holes to obtain electrical connection with the actuating electrodes. Finally, all of the electrodes were covered with insulating ink again except for electric terminals that connect to voltage sources.

The pitch of the actuating electrodes is 1.0 mm with line/space width of 0.7/0.3 mm. Size of each one-DOF electrode measures 48.5 by 43 mm.

V. EXPERIMENTAL EVALUATION OF ACTUATOR

This section evaluates slider motions on the developed 3-DOF electrostatic actuator. A polyethylene sheet with a thickness of 0.1 mm and square size of 90 mm was used as an actuated object. Fig. 7 shows a photo of the sheet and the condition for position detection. The sheet has printed markers with a diameter of 10 mm, at the center and the four corners.

Motion of the actuated sheet was measured using motion analyzing microscope (VM-6000, Keyence Corp.). The frame rate was 250 fps. Applied voltages were generated from waveform generator (7075, Hioki E. E. Corp.) and amplified thousand-fold using high voltage amplifiers (Model 609C-6, TReK Inc.). Applied voltage was three-phase sinusoidal waves with amplitude of 600 V_{0-p}. Glass beads with diameter of 100 μm were scattered between the stator and a slider

 ISBN: 978-1-61208-325-4

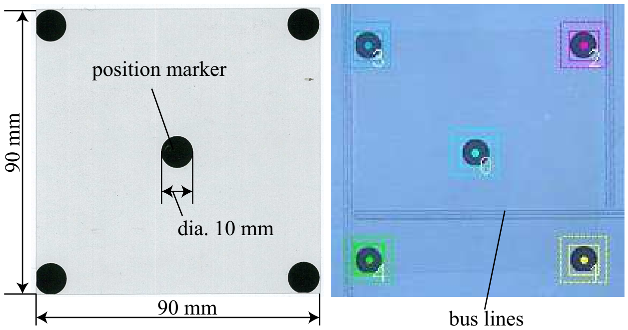


Figure 7: An example of a slider with size of 90mm square used for motion evaluations.

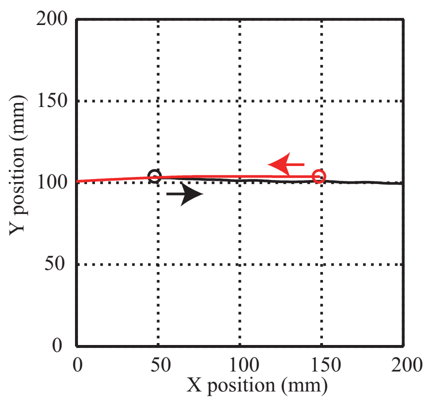


Figure 8: Bidirectional X motion of a 90 square -sized slider.

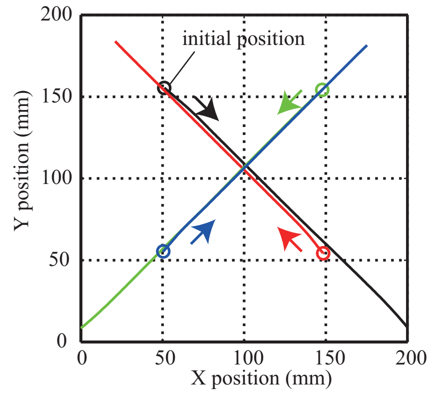
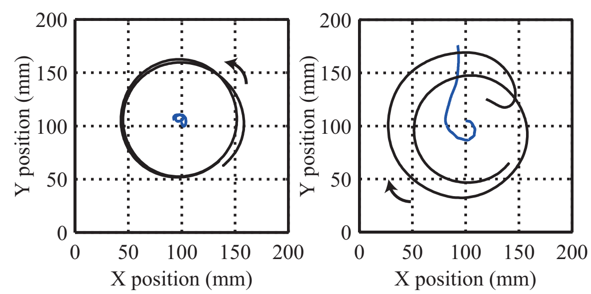


Figure 10: Diagonal motion of a 90 square -sized slider with four directions



(a) stable motion (b) unstable motion

Figure 11: A stable motions and an unstable motion in a rotating point. (a)a stable CCW rotation. (b)an unstable motion

to reduce friction. First, translational motion along X axis was investigated. Changing the applied voltages realized a bidirectional X motion, as shown in Fig. 8. Similarly, Y motions were evaluated using the same slider. Fig. 9 shows the results of Y motions. Diagonal motions were also evaluated by exiting both X and Y electrodes. Trajectories of the slider for the diagonal motion are shown in Fig. 10. These results showed that sliders can be actuated in any directions.

Next, rotating motions were evaluated. The slider was placed on the center of the three-DOF electrode. Two directions of rotation were evaluated, as shown in Fig. 11. Fig. 11(a) showed a stable CCW rotation at the center point. In contrast, the CW motion failed in a short time, as shown in Fig. 11(b). The cause of the failure is outward force direction. When the sheet was rotated in CCW direction, all the four sets of the actuating electrodes generate inward forces, which keep the sheet at the rotation point. On the other hand, when CW rotation was tried, all the four electrode sets produce outward forces, in such a case, the rotation point is unstable equilibrium for the sheet.

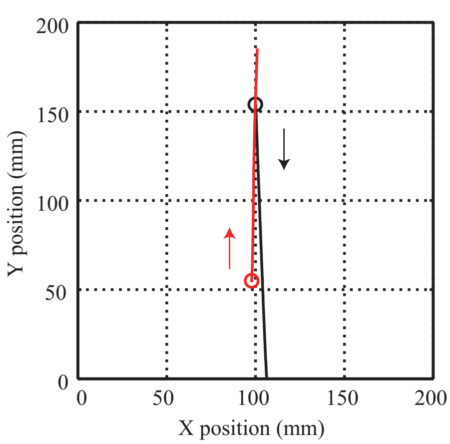


Figure 9: Bidirectional Y motion of a 90 square -sized slider.

Through these results, we confirmed that the newly designed 3-DOF actuator can provide quasi 3-DOF motion, which are X and Y translation and rotation at selected points.

VI. ACTIVE TANGIBLE INTERACTION SYSTEM

A. System overview

Finally, a tangible interaction system was prototyped by placing the 3-DOF actuator on top of LCD of a digitizer pen-tablet. The overview of the prototype system is shown in Fig. 12. We utilized a digitizer (Cintiq 22HD, Wacom Co., Ltd), which has a LCD display of 1920×1080 pixels and position detecting function using electromagnetic induction, which is compatible with the electrostatic actuator.

Two sets of the three-DOF actuators were placed on the display. A polyethylene sheet with a position and orientation sensor was placed on the display as an actuated object. The position and orientation sensor was extracted from an input pen of the digitizer. The circuit was cut and placed in a plastic container, as shown in Fig. 13. The size was $39 \times 16 \times 18$ mm and the weight was 4.6 g. The XY resolution of the digitizer was set as 19200×10800 for the whole area and angle resolution of one degree. The total weight of the actuated

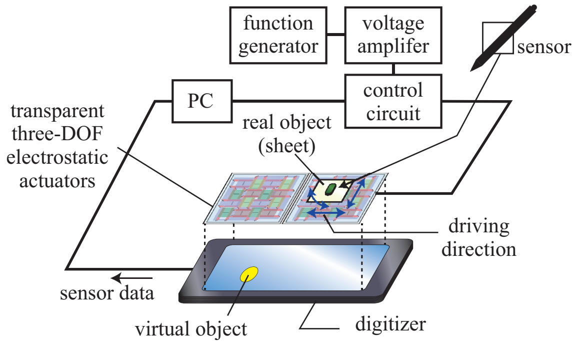


Figure 12: Overview of whole system using two 3-DOF electrodes placed on a digitizer.

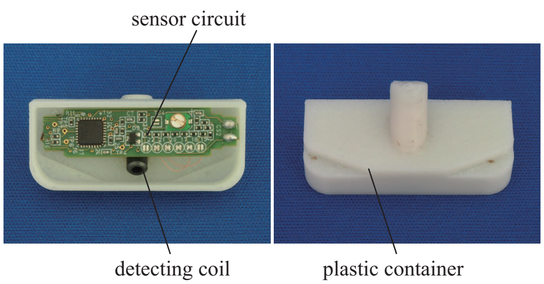


Figure 13: Position and orientation sensor fabricated using a circuit of an input pen of the digitizer.

object including the sensor shown in Fig. 14 was 5.8 g. Glass beads with 100 μm were scattered between a stator an a slider. A three-phase voltage with amplitude of 600 V_{0-p} and frequency of 100 Hz was used for actuation. Voltages were switched on and off by photo-mos relays, depending on the status of the application software and driving directions. Motions of the slider will be asynchronous under the voltage condition.

B. *Position and orientation detecting*

The position and orientation of the slider were detected using the digitizer, as shown in Fig. 15. The square in pink color is a temporary marker to indicate the detected position and orientation. In a relatively slower motion, the position and orientation are successfully obtained, and the marker position and orientation match with the slider, as shown in Fig. 15 (a). In contrast, the digitizer output shows a delay in a faster motion such as more than 100 mm/s, as shown in Fig. 15 (b). Since the program ran at about 100 Hz and the digitizer signal was successfully obtained at each time loop, the delay is not attributed to the program; it would be probably due to internal delay of the digitizer.

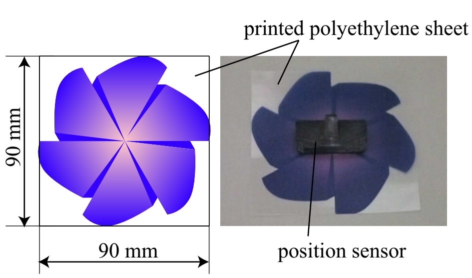
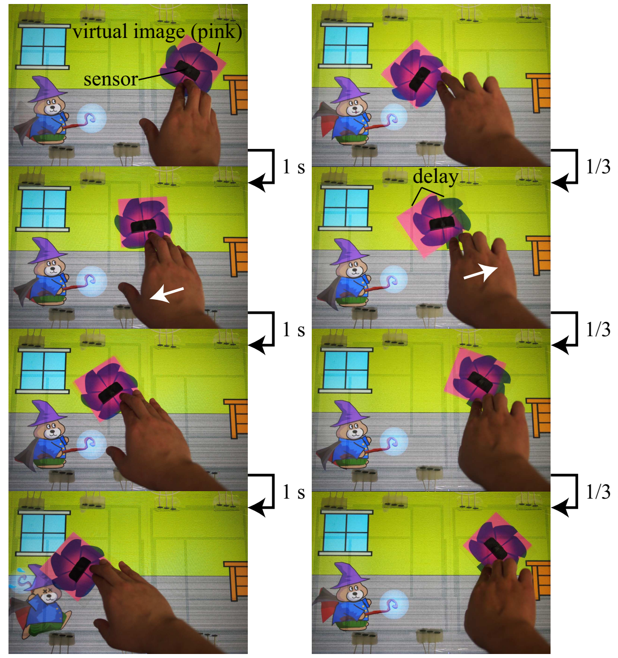


Figure 14: Slider of a polyethylene sheet with a fabricated position sensor.



(a) slower motion (b) faster motion

Figure 15: Detection of position and orientation on the interaction system. (a) In slower motion (b) In faster motion.

C. *Demonstration*

We programmed a simple application using the system, as shown in Fig. 16. A movie of the application can be found on the Internet [11]. In this application, a virtual character representing a wizard is walking around on the screen. Users can interact with the wizard through the actuated sheet. When a user throws the sheet to hit the character, it shows a puzzled behavior and steps back. At the same time, the sheet bounces back on the character. This motion was achieved by applying appropriate voltages to each electrode set based on the position and orientation of the slider. Asynchronous feature showed smooth motions without step-out vibration in cases of changing direction where higher driving forces were required. As another feature, the wizard character can throw a magic ball to the slider. The ball chases the slider and when they contact, the slider starts to rotate, as shown in Fig. 17. The asynchronous driving realized smooth and continuous rotation without any step-out. These results successfully demonstrated availability of a newly designed three-DOF transparent electrostatic actuator for a three-DOF user interaction.

VII. Conclusion and Future Work

This paper introduced a three-DOF transparent electrostatic actuator for active tangible user interaction. Asynchronous driving characteristics of the actuator achieved smooth X, Y, and diagonal translations and rotations of a plastic sheet. The smooth motions are suitable for a dynamic application. Combination of a digitizer and three-DOF actuators achieved three-DOF dynamic interaction between a user and a virtual character using a sheet-type real object. The simple demonstration program confirmed that the newly developed three-DOF transparent electrostatic actuator is suitable for a dynamic

Figure 16: Collision between a virtual character and a real object.

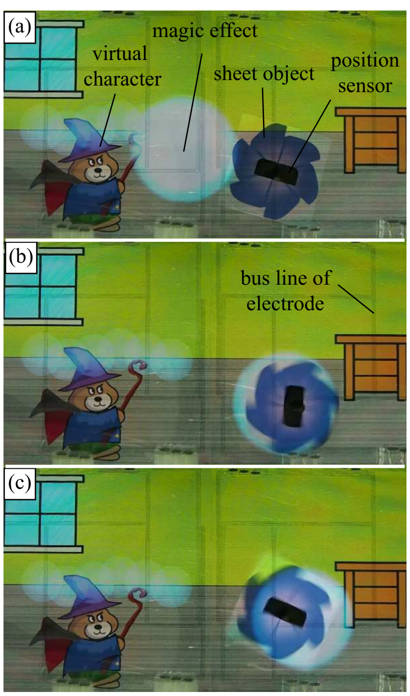


Figure 17: Rotating motion caused by a hit of virtual magic ball.

tangible user interaction. The usability of the proposed system would be evaluated in future work.

ACKNOWLEDGMENT

This work was supported by Funding Program for Next Generation World-Leading Researchers (NEXT) from Japan Society for the Promotion of Science.

REFERENCES

[1] S. Brave, H. Ishii, and A. Dahley, "Tangible Interfaces for Remote Collaboration and Communication", Proc. CSCW'98, Nov. 1998, pp. 169-178.

[2] D. Avrahami, J. O. Wobbrock, and S. Izadi, "Portico: Tangible Interaction on and around a Tablet", Proc. UIST'11, Oct. 2011, pp. 347-356.

[3] G. Pangaro, D. Maynes-Aminzade, and H. Ishii, "The Actuated Workbench: Computer-Controlled Actuation in Tabletop Tangible Interfaces", Proc. UIST'02, Oct. 2002, pp. 181-190.

[4] S. Yoshida, H. Noma, and K. Hosaka, "Proactive Desk II: Development of a New Multi-object Haptic Display Using a Linear Induction Motor", Proc. IEEE Virtual Reality Conference 2006, Mar. 2006, pp. 269-272.

[5] K. Amano and A. Yamamoto, "An Interaction on a Flat Panel Display Using a Planar 1-DOF Electrostatic Actuator", Proc. ITS 2011, Nov. 2011, pp. 258-259.

[6] K. Amano and A. Yamamoto, "Tangible Interactions on a Flat Panel Display Using Actuated Paper Sheets", Proc. ITS 2012, Nov. 2012, pp. 351-354.

[7] T. Hosobata and A. Yamamoto, "Mixed Reality System on Flat Panel Display with Real Object Driven by Synchronous Transparent Electrostatic Actuator", Proc. Int. Conf. Multimedia and Human Computer Interaction, July 2013, pp. 127-1-127-7.

[8] S. Egawa, T. Niino, and T. Higuchi, "Film Actuators: Planar Electrostatic Surface-Drive Actuators", Proc. 1991 IEEE MEMS, Jan.-Feb.1991, pp. 9-14.

[9] T. Hosobata, A. Yamamoto, and T. Higuchi, "2-DOF Synchronous Electrostatic Actuator with Transparent Electrodes Arranged in Checkerboard Patterns", Proc. 2013 IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS), Vol. 1, Nov. 2013, pp. 4919-4924.

[10] A. Yamamoto, S. Tsuruta, and T. Higuchi, "Planar 3-DOF Paper Sheet Manipulation Using Electrostatic Induction", Proc. 2010 IEEE ISIE, July 2010, pp. 493-498.

[11] http://www.youtube.com/watch?v=bht9B5ku5R8 (last access to on 30/Jan./2014)

Design Practice in Human Computer Interaction Design Education

Alma L. Culén
Department of Informatics
University of Oslo
Oslo, Norway
almira@ifi.uio.no

Henry N. Mainsah
Centre for Design Research
School of Architecture and Design
Oslo, Norway
Henry.Mainsah@aho.no

Sisse Finken
Department of Informatics
University of Oslo
Oslo, Norway
finken@ifi.uio.no

Abstract—**As the need for Human Computer Interaction (HCI) designers increases so does the need for courses that best prepare students for their future work life. Multidisciplinary teamwork is what very frequently meets the graduates in their new work situations. Preparing students for such multidisciplinary work through education is not easy to achieve. In this paper, we investigate ways to engage computer science students, majoring in design, use, and interaction (with technology), in design practices through an advanced graduate course in interaction design. Here, we take a closer look at how prior embodied and explicit knowledge of HCI that all of the students have, combined with understanding of design practice through the course, shape them as human-computer interaction designers. We evaluate the results of the effort in terms of increase in creativity, novelty of ideas, body language when engaged in design activities, and in terms of perceptions of how well this course prepared the students for the work practice outside of the university.**

Keywords—HCI education; interaction design; studio; design education; multidisciplinary teamwork.

I. INTRODUCTION

There is an increased movement towards informing and embedding education practices from other disciplines into Human Computer Interaction (HCI). This is especially true when it comes to design practice and design pedagogy [1].

Many authors have stressed a need for considering new pedagogical approaches to HCI education that creatively synthesize HCI theory and methods with design thinking-in-action (see, for example, [2]–[5]). Faiola has argued for development of pedagogical models intended for teaching HCI that *"provide students with knowledge domains that can account for understanding design, social context, and business strategies in addition to computing"*, [6, p. 30].

Winograd and Klemmer, discussing the reasoning behind opening the now famous d.school at Stanford, an innovation hub with a core in human computer interaction design, state: "*The basic premise of the d.school is that students need two complementary kinds of training. The disciplinary training provided by conventional departments provides them with depth in the concepts and experience of a specific field. This gives them intellectual tools, but often misses the larger context of relevance and integration with other kinds of knowledge, which are required to innovate effectively in the 'real world'"*, [7, p. 1]. Such multidisciplinary and effective learning arenas are not easy to create. They represent innovative thinking and innovative education, which has not yet been able to prove itself worthy over time. Thus, embedding innovative educations into traditional educational institutions is difficult. However, the evidence is there that the multidisciplinary approach, such as d.school, has its merits. In line with how Bannon argues why HCI needs to change in the 21st century [8], we argue that the HCI educations needs to change in order to accommodate for new technologies, new interaction forms, new practices, and new areas of research. One practice outside the traditional HCI field, which has a strong influence on changes taking place within HCI, is the design practice. Many scholars have explored the relation between HCI and design. Some of the notable results of these explorations are: a proposition to consider HCI as research through design, see [9]–[14], a proposition to consider Human Computer Interaction Design (HCID) as a radically interdisciplinary dialogue [15], convergent - divergent questioning [16], models, theories and frameworks toward a multidisciplinary science [17].

Two of the authors, of this paper, work within department of informatics, teaching traditional HCI and qualitative research methods. The third author works at a traditional design institution, the school of architecture and design. Over the past few years, the two schools have cooperated and run a graduate course in interaction design together. The course took place at the school of architecture and students from both institutions worked on design projects in multidisciplinary teams. The cooperation came to an end this year, as the design school faculty felt that the differences in traditions and practices between the two schools were too far away from each other. This situation was the immediate motivator for exploring different venues and different approaches to teaching design practice and design thinking within the department of informatics.

In this paper, we present the teaching approach that we have chosen and the results of applying the design pedagogy in the context of an advanced HCID course in the department of informatics. Our goal was not to educate designers, but to teach HCI students about design practice through direct experience and reflection. Similar approaches have been advocated by other scholars, e.g., [3], [10], [18]. Our intention with the course was to prepare HCID students for better collaboration in multidisciplinary teams, to bridge some of the differences in the traditions, to learn, inspired by the d.school, about design thinking, design practice, and to understand a reflexive practice. In order to evaluate the success of our approach we have chosen the following criteria: emergence of creativity, novelty of generated ideas, body language when engaged in design activities and perception of how well the course prepared students for the work practice outside the university.

The paper is structured as follows: in Section II, we briefly describe the HCI pedagogical models that the students were familiar with, and then a design model used by the school of architecture to teach interaction design. We proceed to explaining, in Section III, our case, where the new pedagogical model was applied to teach a course in HCID. In Section IV, we discuss our findings, and sum up the paper in the concluding Section V.

II. Pedagogical Models: HCI and Design

A. HCI model

Many HCI pedagogical approaches include a mix of user-centered requirement analysis, design, implementation and evaluation [19]. This mix is exactly what our students have received through three HCI courses, which they had taken prior to the graduate class described here. Their first course covers material from the book [20]. In the second course, students gain theoretical and practical knowledge on how to study situated use of technology and how such studies can inform design of technology. Reimer and Douglas [21] point out that such study program often falls short of teaching students good design of real-world artifacts, while engaging in real-world design processes. In order to address the real-world settings, the third course in HCI, using [22] as a course book, defines projects based on the needs of local companies and organizations. All three HCI courses are project-based courses. However, they use a classical teaching model consisting of two hour-long lectures, in a lecture hall, and two hour-long sessions in smaller groups. The later provides help with exercises from the book, questions around the material covered during the lectures, or issues related to the project work. The third course in HCI also offers an hour-long design feedback session with the instructor and a representative of a company for which the students are designing. The projects are carried out in project teams of 3-4 students. Although the third HCI course addresses the issue of real-life problems, there is still a gap between multidisciplinary teamwork in professional circles and what students can experience in terms of teamwork in the context of this HCI course.

Another important aspect of learning, present in some design disciplines and often lacking within HCI, is related to approaches that emphasize speculative and inductive ethos. Lewis argues in [23] that technology education nowadays needs to promote more than simply knowledge of materials, mastery of special technical skills and techniques, or correct use of tools or instruments. It should move beyond these to pursue "more subjective and elusive goals", [23, p. 35]. Among these goals he includes creative insight. According to Lewis, the teaching of design is ideally suited to uncover students' creative potentials, because design allows open-endedness [23, p. 45]. Design problems are ill-structured, solutions are not defined in advance, and pathways to the solution are open. Cropley [24] identifies these issues as precisely the conditions that promote creativity. Creativity can be nurtured through a pedagogical framework that builds on open-ended problem solving, using design processes for real-life contexts [25].

While *"HCI specialists still focus on the issues that gave birth to the field: Are technologies learnable, usable, useful, reliable, comprehensible, ethical? We are still concerned with assessing whether technologies serve, engage, and satisfy people and extend their capabilities, or frustrate, thwart, and confound them"*, as the authors state in [26, p. 44], so does most of HCI education as well. In order to answer questions about users and technology, the focus is, naturally on users and what they do with the technology. The students of HCI are thus also trained in seeking the input from users, whether it is for research or design purposes. However, as Bødker [27] points out, the so-called third wave of HCI includes broader consideration of cultural and historical embeddedness of technology, also in non-work contexts, where emotion and aesthetics play a much larger role. The third wave of HCI has, therefore, comes closer to traditional design disciplines, not only through aesthetics, but also focus on solving real-life problems through design of technological solutions.

B. Design model

Hoadley and Cox state: *Design is an important class of human activity because it links theory and practice, bridging scientific activities with creative ones in order to deal with ill-structured, open-ended problems*, [28, p. 20]. To solve problems for real-life contexts, designing combines formal knowledge, experience, practice and judgment, both through and in action. Schön proposes an *"epistemology of practice implicit in the artistic, intuitive processes which some practitioners bring to situations of uncertainty, instability, uniqueness, and value conflict,"* that he characterized as "reflective practice" [29]. Design pedagogy can be understood through a socio-cultural perspective on learning that is centered on developmental aspects occurring between cultural and socially mediated actions in contemporary and legacy contexts [30].

Design studio learning and teaching relies on the integration between people. Schadewitz and Zamenopoulos say: *"The studio model has fostered the type of enculturation into practice that modern schemes for distributed situated learning are just coming to understand,"* [31, p. 1]. Shaffer [32] presents the academic design studio as a coherent system where surface structures, pedagogy, and epistemology interact to create a unique learning community. Surface structures refer to components of the learning environment such as the space, furniture, assignments and so forth. Pedagogical activities include activities such as iterative design cycles, field research, and group discussions of work in progress. Epistemological understanding describes the beliefs and the nature of design knowledge and how it is constructed [3]. Brandt and colleagues draw upon Lave and Wenger's concept [33] of "communities of practice" that describes learning communities where novices are first introduced as legitimate peripheral participants and integrated more centrally into the community through their participation in increasingly more complex tasks. Learning-in-practice is contextual and situated in time and space and is shaped by the historical dimensions of institutions and participants' own life experiences which contribute to shape the manner in which the learning environment is enacted. In this way, it is similar to an "ecological approach" to understanding learning in a more holistic way that brings together a focus on practice, tools, learning environments, and social context; see [3, p. 336]. The teaching rarely involves research articles, books to teach from, or regular formal lectures.

We are now in a position to present our case and describe challenges and lessons learned from introduction of design studio practices into teaching an advanced HCI class.

III. The Case: Design Practices in HCI Teaching

In implementing a design practice in the context of the class we chose to work in the lab, where it was possible to implement practice-based learning. The students had access to materials such as scissors, paint, fabric, paper, tools like sawing machine, hammers, pliers and like, as well as electronic components, such as Lilypads, GPS sensors, LED lights, wires, welding station, etc. The number of students was restricted to ten. They were all advanced graduate students with three prior courses in HCI, as described in the previous section, including experience with user-centered and participatory design approaches. The teaching team consisted of the two in-house teachers and one teacher from the school of architecture and design. The later attended the class approximately every third week, providing feedback on the students' design projects. The authors of this article are the three teachers who have an insatiable curiosity about creativity and how it emerges. Specifically, we are curious about what happens when HCI practitioners cannot rely on the usual ways of thinking and working – that is, when they do not have the support of users in the design process. It is commonly considered that HCI designers should design interactive products *"to support the way people communicate and interact in their everyday and working lives"*, [20, p. 9]. In order to design such products, HCI designers rely on user participation and user studies. These studies inform the design, but also split the responsibility for "good" design between designers and users who informed it. It could be said that users are a great support to HCI designers in, at least, the following ways: helping in testing products and prototypes, informing design processes, participating in them, allowing designers to observe them interacting with technology and last but not least, by making HCI designers feel that designs processes are not dependent on the mystery of creativity and creative processes.

Throughout the course, the in-house teachers have uploaded literature of relevance to a dropbox. This literature covered a range of different subject matters such as: design thinking, design anthropology, differences between interaction design practices within design and HCI, service design, participatory service design, design research, and an article concerning design of wearable technology. Some of the papers were chosen in response to students' project ideas and others aimed at explaining the differences between practices of Interaction Design (ID) and HCID.

Instead of having the traditional lectures, the model described earlier which the students are used to, the shared time between the teachers and the students was spend on discussing various topics, design ideas, and on providing feedback on designs in progress. Some, perhaps unusual forms of stimulating students to be more open and creative were used. For example, in order to increase the energy level and engagement, we would form a circle, and in turn, everyone had to "design" a move that the whole circle then repeated for a while, and then proceeded to another person (see Fig. 1). Other times, we encouraged new ways of exploring the world [34]. For example, we brought artists with interesting ideas and products into the classroom; see Fig. 2. Altering the 'lecture set-up' in this manner was a part of the pedagogical aim of introducing new ways of conducting HCI teaching, with intention to increase the D(esign). In doing this, we aimed at making the students step outside their comfort zone, as well as encourage bodily engagement and hands-on design practices.

Figure 1. Students and a faculty member standing in a circle and "creating" new body movements.

In order to further support the bodily engagement and hands on practices, the students were asked to complete two projects during the semester. The aim of the first project was to design an exhibit addressing the activities and research interest of the group for design. The exhibit was shown at a faire, (which is held annually at the department of informatics), presenting the work of different research groups. The faire also featured representatives from many local IT companies. The second project was to design an installation for the library. The interactive installation had as a goal to bring forward those resources and services, available through the department's library, which usually remain hidden or under-used.

During the first eleven weeks of the course, we carefully documented the students' work on their first design projects.

Figure 2. Amanda Steggell, an artist who made the energy bank, a charging station for mobile phones, showed her product (in the red square, [35]).

In documenting the process we took photographs and collected Post-it notes, which we used to quickly note input (aim, what, how, do-ability) during a feedback session. Further, we took notes during conversations with students, or when they presented their work. Additionally, we handed out briefs with targeted questions concerning creativity, work effort in class, expectancies of outcomes from the course, and addressed the issues around multidisciplinary work. These targeted questions were answered either orally or in writing.

Both in-house teachers and students have taken photographs, amounting to over 300 images, and have shared them in the aforementioned dropbox. The dropbox was made specifically for this class to share photos, presentations of the design projects, and other class related material, such as the literature. The photographs, used in this paper, are from the shared pool in the dropbox. From a teacher's perspective the photographs have been a way of documenting [36] the process from the first drafting of ideas to the materialization of the designs. In addition, the photographs have served as information, beyond mere documentation, and have been used as entrances to gain understandings about increase in creativity, novelty of ideas, and body language when students engaged in design activities; see Fig. 2 and Fig. 4 – Fig. 7.

IV. Discussion

From our experience of cooperating with the school of architecture and design and projects based on multidisciplinary teamwork, we have learned that the building of "communities of practice" did not work for us. The HCID students were not considered to be "real" part of the design team, but rather programmers waiting around for the time to come when technology is embedded into the product, and thus, they were never really integrated into the community of practice. This is, in part, natural, as they never experienced studio work before and needed time to understand surface structures, pedagogical activities and epistemological beliefs. The HCID students' work was never publically criticized before, nor was it ever exhibited for others to see. In addition, they have little practice in speaking about their work. The experience with shifting from humbleness underway to pride in ownership of the work they did and the ideas that led to the final product, was also new.

In our case, as suggested in [3], the students could not be introduced as novices and gradually integrated more centrally into the community of design practice through their participation in increasingly more complex tasks. We had to find a way to bring them closer to understanding and experiencing of that practice on our own. The start of this process; see Fig. 3, was not easy. It was clear that the students used "re-cycled" ideas, and simply changed the domain to which they applied them. For example, finding the faces in places (finding things that resemble faces, Fig. 3 upper right corner) was suggested as an extra activity during the exhibit. Never mind the fact that none of us, faculty or students, could find faces in the building, making the task really hard and little fun as an activity on the day of the exhibit, the idea itself was not novel and neither was a way of using it [34].

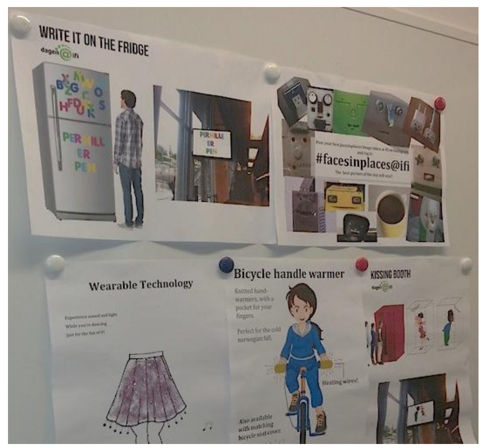


Figure 3. Among the initial ideas one can see recycled ones, such as the "faces in places", mixed in new ones, such as heated bicycle gloves for those who harsh nordic winters do not hinder from using the bike.

Figure 4. The body language of those present during the first feedback session shows little excitement or passion.

Students' (and faculty's, far right and far left on Fig. 4) body language also shows that there was little entusiasm for those ideas and concepts, most of which were based on pure entertainment.

Figure 5. The process of making things for the exhibit.

Critiques, especially in the beginning, were often taken personally. However, making things, such as the skirts or cushions for the "iConfession booth", slowly placed smiles on students faces; Fig. 5. It really helped to start unfolding some of internal processes leading to increased creativity and willingness to learn new skills or use the existing ones in order to further the processes that the group was engaged in. We have reported on the emergence of creativity in this context, and used assemblages of skills as a framework to analyze it [37].

Figure 6. The students working on the exhibit site, the exhibits representative of research on sustainable design, privacy and wearables.

After the first six weeks, there was a breakthrough. The first project was organized as exhibit consisting of three parts. The first part built on the idea of sustainability, producing the energy while biking, to power blinkers built into the glove, as well as to heat gloves, light up the wheels, etc. The presentations now included sketches, material choices, hand knitted gloves; see Fig.7.

Figure 7. Sketches of the prototype on the left, palette of material choices, as well as very handmade gloves, connected to a small dynamo.

The second part was related to design group's research projects regarding design for people with dementia. The skirt for dement ladies was the result, as as a counter-balance a "bliky" party skirt that lights up when the proximity sensor is activated; see Fig. 8 and [38], [39].

Figure 8. On the left, the skirt with a proximity sensor. Remaining images are of the skirt for dementia: comfort balls, the GPS and a QR code.

The third part was an iConfession booth, a tool for exploring the anonymity and willingness of people to disclose a secret; see Fig. 6 and [40], [41].

Finally, all three parts were put together; Fig. 6 and Fig. 9 show making of the exhibit and the final result, respectivelly.

The ideas implemented, although inspired by one thing or another, were novel. The audience received the exhibit very well, and students have expereinced the sense of pride and satisfaction with the result.

In order to get feedback directly from students around their perception of readiness to partake in multidisciplinary teams, we have designed a short questionnaire. Asked about the difference between the usual HCI classes and this one, they said: *The course focused more on design thinking, and to get a finished product to exhibit. It was less literature, testing, and report writing than other HCI courses.* All students expressed that they have better understanding of the design practice. In a short survey after the course we asked them if they thought that articles were helpful; half of the students answered that they thought so, while the other half thought that for this course the articles were not so important. When asked if they thought that design oriented practices (making things), have given them new skills, as HCID designers, six students agreed strongly, one did not reply and three were neutral. As for the perception of how well they are prepared after the course for work with designers, six students answered that they strongly feel that they are better positioned for such cooperation, and four did not have strong opinion about the issue, but were not negative.

Our understanding is that the experience they gained may help them discover who they are as HCID designers [42, 43], both by understanding the difference between the HCI practice and the design practice, and by direct experience of the design practice.

V. CONCLUSION AND FUTURE WORK

The lessons learned and discussed in this paper show that HCID students could adopt and understand design practices, in spite of a rather long experience and a strong sense of being rooted in the HCI tradition. The teachers, the students, and the audience at the exhibit have all been satisfied with the exhibit in terms of adequateness of concepts in relation to design task, prototypes developed and organization of those into an exhibit. The students' body language has changed from indifferent and closed to engaged and open. They all perceive this piece of learning to prepare them for the work as professional interaction designers better than the HCI courses alone could do. Thus, we conclude that this approach warrants further exploration. As future work, we would like to follow these students into their work life and see if this experience had an impact when working in teams with designers in real-life projects.

ACKNOWLEDGMENT

Thanks go to all of our students: Andrea Gasparini, Henrik Kjersem, Lena Risvik, Paria Tahaee, Anja Simonsen, Sylvia Saxlund, Rita Johnsen, Ingrid Arnesen, Agnethe Heggelund and Rebekka Castro, and our guest lecturers Tom Igoe, Amanda Steggell and Lavrans Løvlie.

REFERENCES

[1] P. Koutsabasis and S. Vosinakis, "Rethinking HCI education for design: problem-based learning and virtual worlds at an HCI design studio," International Journal of Human-Computer Interaction, vol. 28, no. 8, 2012, pp. 485–499.

[2] E. Blevis, "Design challenge based learning (DCBL) and Sustainable Pedagogical Practice," interactions, vol. 17, no. 3, 2010, pp. 64–69.

[3] C. B. Brandt, K. Cennamo, S. Douglas, M. Vernon, M. McGrath, and Y. Reimer, "A theoretical framework for the studio as a learning environment," Int. Journal of Technology Design Education, vol. 23, no. 2, 2013, pp. 329–348.

[4] J. Kolko, Thoughts on Interaction Design, 1st edition. Morgan Kaufmann, 2010.

[5] S. E. Lundgren, O. Torgersson, L. Hallnäs, P. Ljungstrand, and E. Erikson, "Teaching interaction design: matters, materials and means", DRS Wonderground International conference, IADE, 2006, pp. 1–4.

[6] A. Faiola, "The design enterprise: rethinking the HCI education paradigm," Design Issues, vol. 23, no. 3, 2007, pp. 30–45.

[7] T. Winograd and S. Klemmer, "HCI at Stanford University," Interactions, vol. 12, no. 5, 2005, pp. 30–31.

[8] L. Bannon, "Reimagining HCI: toward a more human-centered perspective," Interactions, vol. 18, no. 4, 2011, pp. 50–57.

[9] D. Fallman, "Design-oriented human-computer interaction," in Proceedings of the SIGCHI conference on Human factors in computing systems, 2003, pp. 225–232.

[10] N. True, J. Peeters, and D. Fallman, "Confabulation in the time of transdisciplinarity: reflection on HCI education and a call for conversation," in Human-Computer Interaction. Human-Centred Design Approaches, Methods, Tools, and Environments, M. Kurosu, Ed. Springer Berlin Heidelberg, 2013, pp. 128–136.

[11] J. Forlizzi, J. Zimmerman, and S. Evenson, "Crafting a place for interaction design research in HCI," Design Issues, vol. 24, no. 3, 2008, pp. 19–29.

[12] J. Zimmerman, J. Forlizzi, and S. Evenson, "Research through design as a method for interaction design research in HCI", Proceedings of the SIGCHI conference on Human factors in computing systems, 2007, pp. 493–502.

[13] J. Zimmerman, E. Stolterman, and J. Forlizzi, "An analysis and critique of research through design: towards a formalization of a research approach," in Proceedings of the 8th ACM Conference on Designing Interactive Systems, 2010, pp. 310–319.

[14] E. Goodman, E. Stolterman, and R. Wakkary, "Understanding interaction design practices," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011, pp. 1061–1070.

[15] P. Wright, M. Blythe, and J. McCarthy, "User experience and the idea of design in HCI," in Interactive Systems. Design, Specification, and Verification, S. W. Gilroy and M. D. Harrison, Eds. Springer Berlin Heidelberg, 2006, pp. 1–14.

[16] C. L. Dym, A. M. Agogino, O. Eris, D. D. Frey, and L. J. Leifer, "Engineering design thinking, teaching, and learning," Journal of Engineering Education, vol. 94, no. 1, 2005, pp. 103–120.

[17] J. M. Carroll, HCI models, theories, and frameworks: toward a multidisciplinary science. Morgan Kaufmann, 2003.

[18] K. Cennamo, S. A. Douglas, M. Vernon, C. Brandt, B. Scott, Y. Reimer, and M. McGrath, "Promoting creativity in the design studio," Proceedings of the 42nd ACM technical symposium on Computer science education, 2011, pp. 649–654.

[19] S. Greenberg, "Embedding a design studio course in a conventional computer science program," in Creativity and HCI: From Experience to Design in Education, P. Kotzé, W. Wong, J. Jorge, A. Dix, and P. A. Silva, Eds. Springer US, 2009, pp. 23–41.

[20] H. Sharp, Y. Rogers, and J. Preece, Interaction Design: beyond human-computer interaction, 2nd ed. Wiley, 2007.

[21] Y. J. Reimer and S. A. Douglas, "Teaching HCI design with the studio approach," *Computer science education*, vol. 13, no. 3, pp. 191–205, 2003.

[22] D. J. Lazar, D. J. H. Feng, and D. H. Hochheiser, Research methods in human-computer interaction. John Wiley & Sons, 2010.

[23] T. Lewis, "Coming to terms with engineering design as content," Journal of Technology Education, vol. 16, no. 2, 2005, pp. 34–51.

[24] D. Cropley and A. Cropley, Engineering creativity: a systems concept of functional creativity. Lawrence Erlbaum Associates Publishers, 2005.

[25] A. M. Hill, "Problem solving in real-life contexts: an alternative for design in technology education," International journal of technology and design education, vol. 8, no. 3, 1998, pp. 203–220.

[26] E. F. Churchill, A. Bowser, and J. Preece, "Teaching and learning human-computer interaction: past, present, and future," Interactions, vol. 20, no. 2, 2013, pp. 44–53.

[27] S. Bødker, “When second wave HCI meets third wave challenges,” Proceedings of the 4th Nordic conference on Human-computer interaction: Changing roles, 2006, pp. 1–8.

[28] C. Hoadley and C. Cox, “What is design knowledge and how do we teach it?” Educating learning technology designers: Guiding and inspiring creators of innovative educational tools, 2009, pp. 19–35.

[29] D. A. Schön, The reflective practitioner: How professionals think in action, vol. 5126. Basic books, 1983.

[30] J. V. Wertsch, Mind as action. Oxford University Press, 1998.

[31] N. Schadewitz and T. Zamenopoulos, “Towards an online design studio: a study of social networking in design distance learning,” International Association of Societies of Design Research Conference, IASDR, 2009, pp. 1–11.

[32] D. W. Shaffer, “Learning in design,” in Foundations for the Future in Mathematic Education, R. A. Lesh, E. Hamilton and J. Kaput, Eds. Routledge, 2007.

[33] J. Lave and E. Wenger, Situated learning: Legitimate peripheral participation. Cambridge university press, 1991.

[34] K. Smith, How to be an explorer of the world: portable life museum. New York: Perigee, 2008.

[35] “Energy Bank.” http://www.energy-bank.org/. Accessed: January 15, 2014.

[36] M. Crang and I. Cook, Doing ethnographies. Los Angeles; London: SAGE, 2007.

[37] S. Finken, A. Culén and A. Gasparini, “Nurturing creativity: assemblages in HCI design practices”, in Proceedings of DRS 2014, The Big Debates, in press.

[38] R. Johnsen, S. Saxlund, and A. Simonsen, “Skirts with meaning,” skirtswithmeaning, 2013. http://skirtswithmeaning.wordpress.com/. Accessed: January 15, 2014.

[39] A. L. Culén and S. Finken, “A Skirt for Well Aged Ladies with Cognitive Loss,” in Proceedings of HCII 2014, in press.

[40] A. Gasparini, R. Castro, L. Risvik, and A. Heggelund, “iConfess at dagen@ifi,” 2013. http://folk.uio.no/andrega/iConfess/. Accessed: January 15, 2014.

[41] A. L. Culén, S. Finken, and A. A. Gasparini, “iCONFESS: Mirroring Digital Self-Disclosure in a Physical Booth,” in HCII 2014, in press.

[42] A. L. Culén, S. Joshi, and A. Atif, “HCID: who is an interaction designer?” Proceedings of the 2nd International Conference for Design Education Researchers, Oslo, Norway, 2013, vol. 4, pp. 1924–1937.

[43] A. L. Culén and M. Kriger, “HCI in IT-facilitated Business Innovation: a Design Thinking Perspective,” in Proceedings of HCII 2014, in press.

Figure 9. The exhibit addresses sustainable design, anonymity and meaningful wearables. Students recruited participants for the part of the exhibit wearing Guy Fowler masks, a symbol of anonymity, as part of the exhibit activity.

Role of Student Interaction Interface in Web-Based Distance Learning

Marina Mladenova
University of Forestry
Sofia, Bulgaria
e-mail: marina_mladenova@mail.bg

Diana Kirkova
University of Forestry
Sofia, Bulgaria
e-mail: djenidj@yahoo.com

***Abstract* - This article presents a subject to debate the question of the role of learner-interface interaction in distance education. This role is considered in the aspect of possibilities of distance learning in the three main areas of human perception: cognitive, psychophysical and emotional. The results are of the first phase of a study conducted by the authors on the impact of learner-interface interaction in web-based training on the result achieved, including training courses in ICT students at the University of Forestry - Sofia.**

Keywords - web-based distance education; learner-interface interaction; interactive communication

I. INTRODUCTION

Traditional forms of education meet the needs of young people at a time when it was possible to envisage with relative security the knowledge and skills that they would need in their adult lives. Today this is not so. Young people can no longer expect to spend their entire lives in the same economic sector or even on the same place; their career paths will change in a way that cannot be foreseen and they will need a wide range of generic skills that will allow them to adapt. In a more increasingly complex and globalized world, creativity, the ability to think versatile, skills into different areas and adaptability as a trend are valued more than specific knowledge. This led to the development of the concept of education and becoming more widely available web-based distance learning courses in educational institutions. Offering e-learning courses is important for the survival of the modern educational institution [8].

Web-based distance education cannot be implemented without the use of the capabilities of modern information and communication technologies. This is reflected in the adoption in 2010 by the European Council Strategy "Europe 2020" [5], which focuses on active and effective use of the opportunities of modern information and communication technologies for the realization of the idea of widely available, tailored to individual needs, continuing lifelong quality education, which all are given equal opportunities to acquire the knowledge and skills necessary for successful social and professional realization.

The purpose of this paper is to suggest the issue of discussion about the role of learner-interface interaction in distance education. In our opinion, this interaction is important for the improvement of personal achievements in the training of individuals, and hence becomes significant for the quality of the education received. The article presents a theoretical framework, a methodology of research and a discussion on the submitted results.

II. THEORETICAL FRAMEWORK AND METHODOLOGY OF RESEARCH

The analysis is based on the adoption of the following theoretical framework:

- Keegan [2] defines distance education as a system characterized by: 1) the separation of the teacher from the learner through a large part of the learning process, 2) the impact of the educational organization, 3) provide an opportunity to assess and self-assess of the learners, 4) the use of educational media for the presentation of course content, and 5) two-way communication between teacher and student [2].
- Hillman, Willis and Gunawardena [1] added learner-interface interaction as an integral part of the concept of distance learning. This is the interaction that occurs between the learner and the technology used to provide educational content. They claim that the skills of the learner to handle with communication environment is a necessary condition to participate in distance learning course and is positively correlated with success in this course. To acquire knowledge of the course content, the student must be proficient in the technological environment for communication and interaction [1].
- According to Belanger and Jordan [3] learning objectives relate to three main areas of human perception: cognitive, psychophysical and emotional.
- We considered Gestalt psychology (meaning integrity, shape, configuration, connectivity components, closeness, etc.), established in the early decades of the 20th century by Max Wertheimer [9], who studies perception as a single image and the characteristics of the education.

Web-based distance education is learning that takes place via the Internet or Intranet. A Web-browser is mostly used in this type of training, such as Internet Explorer or Firefox. Each training program that uses the Internet as a technology for delivering learning materials, even e-mail correspondence and transfer of files belonging to this type of distance education [11].

Let us consider in more detail the possibilities of distance learning in the three main areas of human perception.

The cognitive domain is divided into six main categories of learning objectives: knowledge, comprehension, application, analysis, synthesis and evaluation. Each category is based on the requirements and characteristics of the category before it, and each category consists of sub categories. Tasks in the cognitive domain can be achieved relatively easily, using technology and distance education through interaction student-learning content.

Psychophysical area focuses on physical skills, manipulation of objects or other physical activities that require muscle coordination. This area is best expressed for example in speech, physical condition, operating machinery or technical courses, which require a physical activity. Multimedia technologies are successfully used to train physical skills - such as aircraft simulators make people learn from their mistakes, which in a real environment may cause a lot.

Emotional domain is focused on individual preferences for perception and processing of information. As Confucius said, "I hear and I forget. I see and I remember. I do and I understand". Here is the key role of the interactions between the participants in the educational process - teachers, learners, learning content and interface.

Today, the majority of platforms for web-based distance education for communication, assessment and self-assessment are set tools that are widely used in social networking as "Blogs", "Journal", "discussion board" and others. This is in order to facilitate perception of the interface by the learner. However, it must be taken into account that the social networks, although popular, are used by less than half of the population.

Interactive communication, which is basically a web-based distance education involves the creation of a dynamic area, changing space, specific navigation, depending on the preferences of the learner, dynamic design, use of specific computer code into high level language, mobile technologies and constant innovations [10]. Things at each visit of the user are different, depending on his skills gathered information on the system and knowledge of the interests and orientation. Three categories of interactivity are distinguished:

- The first category is a navigation interactivity that allows the student to move through the pages of learning content using the appropriate hyperlinks.
- The second category of functional interactivity allows the student to interact with other users, and receive news, are used especially discussion forums, blog and e-mail.
- Third category adapted interactivity allows customization of the site and the browser, depending on the specific consumer preferences.

If the difficulty in interaction student-interface loses the benefits of interactive communication - generating knowledge through exchange of ideas and use of a number of modern, digital media and technologies for data transmission and combination of electronic processes [7]. Interactive communication ensures active participation of those students who do not normally express their ideas openly [6].

The learner should understand and be able to work with the technologies used in the training course. When there is incomprehension of the technological realization of a web-based distance education and the inability of the student to work in such an environment, learner-interface interaction is difficult. As a result, only navigation interactivity is most often used. When the interface impedes the learner achieves the best case, only learner-content interaction.

According to the Gestalt psychology in the field of consciousness there are two contexts - figure and background, and the figure is always clearer, cleaner, more detached, more obvious are its limits. The background looks like continuation of the figure but is darker, more distant, less supportive the figure. Under the law for centering the mind is centered on a figure work with it and focused on it. It is not possible for the entire contents of consciousness, to accept the role of figure.

When the interface impedes the learner, he assumes the role of the figure, the training itself becomes background. Essential educational elements of distance learning as interaction with all members of the school community (student-teacher among themselves students), initiating creative activity manifestation of the student, the active construction of knowledge by means of two-way communication student and teacher are lost. A portion of set in the syllabi of the courses skills and habits cannot be developed by the learner. In this case, a small portion of the planned learning key skills and habits are acquired.

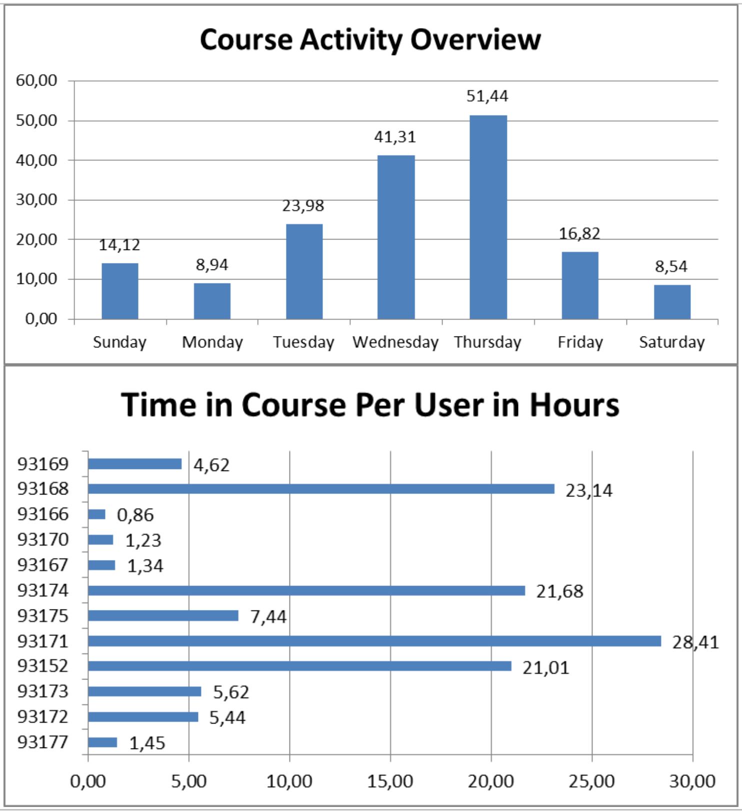
III. RESULTS AND DISCUSSION

The research was conducted for the student interaction interface in web-based distance learning on the result achieved, among training courses in ICT students in University of Forestry – Sofia, during the period October 2013 - February 2014; the web-based platform Blackboard Learn [4] was used for the training. Both training courses include tools for interactive communication and conditionally display the following topics in the structure.

The study was conducted in two target groups: students first course degree “Bachelor” and “Master degree” students.

Activity of students in the "Master degree" is shown in Fig. 1 and activity of students in "Bachelor degree" in Fig. 2.

 ISBN: 978-1-61208-325-4



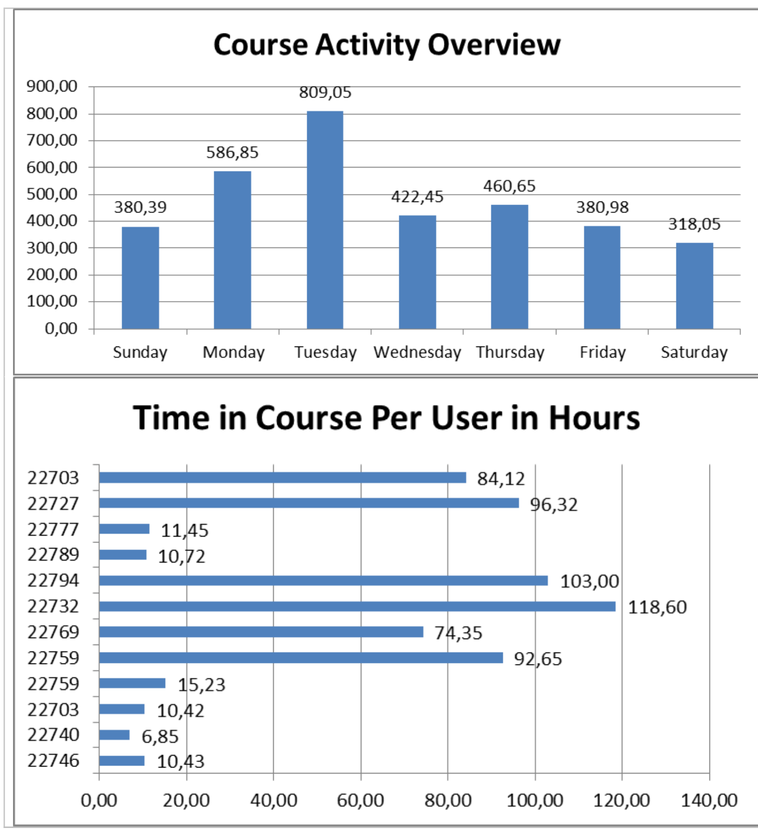
Course Name	**Informatics MD (FELA)**
Course ID	33_19
Number of Students	12
Number of Users	13
Date Range	10/01/2013 - 02/21/2014

Activity is shown for enrolled users only.
Avg Time in Course Per User in Hours – 11,01
Total Time in Course – 165,15

User Activity By Day

10/01/2013 - 02/21/2014

Figure 1. Activity of students in the "Master degree"



Course Name	**Information Technology, BD, FWI**
Course ID	33_14
Number of Students	116
Number of Users	117
Date Range	10/01/2013 - 02/21/2014

Activity is shown for enrolled users only.
Avg Time in Course Per User in Hours – 28.95
Total Time in Course - 3358,42

User Activity By Day

10/01/2013 - 02/21/2014

Figure 2. Activities of students in the "Bachelor degree"

Students in "Bachelor degree" browsed educational content significantly longer than students in "Master degree". This can be explained by the fact that they are 1-st course and most of them do not have a system for the utilization of web-based interactive learning content presented. At the end of the course students responded to the questionnaire survey for the work with the system Blackboard Learn. The following section presents the questions and the results.

1. Did you meet any difficulties working with the platform for customizing your account:

	Bachelor	Master
Strongly Agree	8,62%	8,33%
Agree	12,93%	8,33%
Undecided	8,62%	8,33%
Disagree	32,76%	25,00%
Strongly Disagree	37,07%	50,00%

2. Did you meet any difficulties working with the platform for customizing the system:

	Bachelor	Master
Strongly Agree	15,52%	16,67%
Agree	7,76%	0,00%
Undecided	10,34%	16,67%
Disagree	27,59%	25,00%
Strongly Disagree	38,79%	41,67%

3. Did you meet any difficulties working with the platform for customization of the navigation of the educational content:

	Bachelor	Master
Strongly Agree	26,72%	8,33%
Agree	20,69%	8,33%
Undecided	6,03%	16,67%
Disagree	20,69%	25,00%
Strongly Disagree	25,86%	41,67%

4. Did you meet any difficulties working with the platform while managing the educational content – type file management:

	Bachelor	Master
Strongly Agree	7,76%	0,00%
Agree	6,03%	8,33%
Undecided	1,72%	8,33%
Disagree	37,07%	16,67%
Strongly Disagree	43,97%	66,67%

5. Did you meet any difficulties working with the platform while managing the educational content – type URL:

	Bachelor	Master
Strongly Agree	7,76%	8,33%
Agree	9,48%	0,00%
Undecided	7,76%	8,33%
Disagree	31,03%	41,67%
Strongly Disagree	40,52%	41,67%

6. Did you meet any difficulties working with the platform while managing the educational content – type interactive tutorial:

	Bachelor	Master
Strongly Agree	24,14%	16,67%
Agree	16,38%	0,00%
Undecided	1,72%	0,00%
Disagree	20,69%	33,33%
Strongly Disagree	37,07%	50,00%

7. Did you meet any difficulties working with the platform while managing the educational content – type video:

	Bachelor	Master
Strongly Agree	18,10%	0,00%
Agree	7,76%	0,00%
Undecided	6,90%	8,33%
Disagree	27,59%	25,00%
Strongly Disagree	39,66%	66,67%

8. Did you meet any difficulties working with the platform while managing the educational content – type self test and evaluation:

	Bachelor	Master
Strongly Agree	22,41%	0,00%
Agree	12,93%	8,33%
Undecided	1,72%	0,00%
Disagree	30,17%	33,33%
Strongly Disagree	32,76%	58,33%

9. Did you meet any difficulties working with the platform while managing the educational content – type task:

	Bachelor	Master
Strongly Agree	41,38%	8,33%
Agree	22,41%	0,00%
Undecided	2,59%	0,00%
Disagree	15,52%	33,33%
Strongly Disagree	18,10%	58,33%

10. Did you meet any difficulties working with the platform during the interaction process with the professor:

	Bachelor	Master
Strongly Agree	15,52%	0,00%
Agree	7,76%	0,00%
Undecided	22,41%	0,00%
Disagree	24,14%	58,33%
Strongly Disagree	30,17%	41,67%

11. Did you meet any difficulties working with the platform during the interaction process with the rest of the users during the individual working process.

	Bachelor	Master
Strongly Agree	20,69%	8,33%
Agree	13,79%	0,00%
Undecided	8,62%	0,00%
Disagree	25,00%	41,67%
Strongly Disagree	31,90%	50,00%

12. Did you meet any difficulties working with the platform during the interaction process with the rest of the users during the group working process:

	Bachelor	Master
Strongly Agree	29,31%	0,00%
Agree	18,10%	8,33%
Undecided	6,90%	0,00%
Disagree	25,00%	33,33%
Strongly Disagree	20,69%	58,33%

13. Did you meet any difficulties working with the platform while receiving news:

	Bachelor	Master
Strongly Agree	7,76%	0,00%
Agree	5,17%	0,00%
Undecided	12,93%	0,00%
Disagree	31,03%	25,00%
Strongly Disagree	43,10%	75,00%

14. I met difficulties working with the platform during the process of receiving the actual information about scores:

	Bachelor	Master
Strongly Agree	8,62%	0,00%
Agree	12,93%	0,00%
Undecided	11,21%	0,00%
Disagree	28,45%	16,67%
Strongly Disagree	38,79%	83,33%

15. I met difficulties working with the platform during the process of receiving the actual information about incoming events:

	Bachelor	Master
Strongly Agree	19,83%	0,00%
Agree	5,17%	0,00%
Undecided	7,76%	8,33%
Disagree	30,17%	16,67%
Strongly Disagree	37,07%	75,00%

16. The difficulties in working with the platform do you think is mostly due to:
    a) the platform interface;
    b) personal Internet access;
    c) insufficient skills to work with ICT;
    d) I did not encounter any difficulties.

	Bachelor	Master
the platform interface	37,07%	16,67%
personal Internet access	18,10%	8,33%
insufficient skills to work with ICT	20,69%	8,33%
I did not encounter any difficulties	24,14%	66,67%

17. How do you assess the impact of the use of the platform on the end result of your training and knowledge obtained:

	Bachelor	Master
Excellent	18,97%	16,67%
Good	26,72%	58,33%
Poor	25,00%	8,33%
Very Poor	11,21%	8,33%
No impact	18,10%	8,33%

Respondents are 25% of the students first degree "Bachelor’s degree and "Master’s" (a representative sample of the students in the UF). The results allow the following conclusions:

- Students in degree "Master" encountered significantly less difficulty in learner-interface interaction, regardless of the type of learning resources. This allowed them to concentrate on learning of content and to achieve the final result with less time working in the system;
- Both target groups haven't encountered serious difficulties in customizing their system accounts, the system in general, the access to the learning content type - file, URL, video, receiving news and updated information for the scores;
- Significant differences for the difficulties encountered when working with assessment tools and interaction with other users;
- The reason for the difficulties encountered and the two target groups mainly indicate the platform interface, only then insufficient skills to work with ICT;
- Both target groups assess the impact of the use of the platform on the end result of the training and the knowledge obtained as a very strong and highly.

IV. CONCLUSION AND FUTURE WORKS

In deciding on which technology to use for the realization of distance web-based training, must be given to the ease of use of this technology, i.e., whether users will be able to deal with it.

Time of use of the platform for web-based education is not proportional to the achieved learning outcomes.

The lack of difficulty in learner-interface interaction plays an important role in improving personal achievements in teaching individual learners. Achieved learning outcomes in this case affected only by the personal characteristics of the learner.

When properly structured and built, web-based learning content interaction learner-interface strongly influences the outcome of the training and quality of education.

The study will continue by including students trained in the summer semester of the academic 2013/2014 year and will be repeated in the same target groups during the next two academic years. We hope that after its completion, we will be able to outline clear and definite trends, and the impact of learner-interface interaction in web-based distance learning on the outcome of the training and quality of education.

REFERENCES

[1] D. C. Hillman, D. J. Willis, and C. N. Gunawardena, Learner-interface interaction in distance education: An extension of contemporary models and strategies for practitioners. The American Journal of Distance Education, vol. 8(2), pp. 30-41, 1994[1]

[2] D. Keegan, Foundations of distance education, London: Routledge, 1990[2]

[3] F. Belanger, D. Jordan, Evaluation and Implementation of Distance Learning: Technologies, Tools and Techniques, Hershey, PA: Idea Group, pp.246, 2000[10]

[4] http://elearn.ltu.bg/[9]

[5] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM: 2010:2020:FIN:EN:PDF[8]

[6] Inayat, R. ul Aminb, Z. Inayatc, Siti Salwah Salima Z Inayat, and S. Salim, Effects of Collaborative Web Based Vocational Education and Training (VET) on Learning Outcomes, Elsevier, Computers & Education, Vol. 68, October 2013, pp. 153–166[3]

[7] J. Charles, Interactions in Online Education: Implications for Theory and Practice, Routledge, NY, 2006[5]

[8] J. Kambutu, Administrators prefer technology-based distance learning. The Quarterly Review of Distance Education, 3(3), 341-343, 2002[4]

[9] M. Wertheimer, Gestalt Theory. A source book of Gestalt psychology, London: K. Paul, Trench, Trubner & Co., Ltd., pp. 403–410, 1938[11]

[10] O. Simpson, Supporting students in online, open & distance learning, Routledge, UK, 2013[6]

[11] S. B. Dias, J. A. Diniz, and L. J. Hadjileontiadis, E-Learning Exequibility in the Information and Knowledge Society, Springer, Towards an Intelligent Learning Management System Under Blended Learning Intelligent Systems Reference Library Vol. 59, pp. 3-19, 2014[7]

Experimental Study into the Time Taken to Understand Words when Reading Japanese Sign Language

Mina Terauchi
Polytechnic University
(Visiting Fellow of Kogakuin University)
Kodaira-shi, Tokyo, Japan
e-mail: terauchi@uitec.ac.jp

Keiko Watanabe Yuji Nagashima
Kogakuin University
Hachioji-shi, Tokyo, Japan
e-mail: ed13001@ns.kogakuin.ac.jp e-mail: nagasima@cc.kogakuin.ac.jp

***Abstract*— We are conducting research in areas such as linguistic and cognitive analysis of sign language, and animation, with the aim of assisting the hearing impaired with communication. It was already assumed that hearing-impaired people recognize the meaning of sign language words while the sign language movement is being performed. Based on the results of dialog-based analysis of sign language however, we predicted that hearing-impaired people would understand the meaning of sign language words during the "transition" stages. For the purposes of this paper, we compiled a series of experiment sentences and conducted an experiment to determine the timing with which hearing-impaired people recognized target words in each sentence. The results of the experiment indicated that hearing-impaired people recognized a high percentage of words during the "in-transition" stage. Based on these results, we can assume that hearing-impaired people understand sign language sentences by effectively utilizing hand shapes, movements and information such as expressions and intonation. The challenge for the future is how to harness this information to improve the comprehension abilities of students undergoing sign language education.**

Keywords- Sign Language, predict recognition, transition.

I. INTRODUCTION

We are in the process of conducting linguistic and cognitive research into sign language in order to assist the hearing impaired with communication.

For the purposes of this paper, we conducted an experimental study into the time taken for deaf people to predict and understand words at the ends of sentences when reading Japanese sign language. We divided the sign language sentences used in the reading experiment into three different types depending on the nature of the word at the end of each sentence (the "target word"). Type A sentences contained target words that enabled the reader to correctly interpret the meaning of the sentence. Type B sentences contained target words that were of the correct type syntactically, but that did not enable the reader to correctly interpret the meaning of the sentence. Type C sentences contained target words consisting of fabricated movements based on phonemes that are absent from Japanese sign language ("pseudowords"), making it impossible for the reader to interpret their meaning. We compiled these three types of sentences with assistance from a native signer. We filmed the sentences being signed and then edited the sign language clips for use in the experiment.

For the experiment itself, we asked 13 native signers to read sign language sentences following each of the three patterns. The results were as follows.

II. THE TEMPORAL STRUCTURE OF SIGN LANGUAGE

Sign language consists of two different types of signal; manual signals and non-manual signals. It is a visual language that uses both types of signal simultaneously and continuously. Manual signals (MS) are determined by handshapes, the direction of the palms, the position of the hands and other broad movements, and are used to form words. Non-manual signals (NMS) meanwhile consist of elements such as the shape of the mouth, the signer's eye line, expressions and nodding, and are used for semantic and syntactic purposes.

The temporal structure of sign language can be broadly divided into two categories; "signs" and "transitions". A "sign" starts when the signer has decided and formed the shape for a single sign language word. When the signer has finished moving, the sign then ends the moment before they release their handshape. A "transition" starts when the signer has finished the movements required for a single sign language word, and lasts until the moment before the signer starts to sign the next sign language word, as they move into position. Transitions can be further divided into two sub-categories; "in-transitions" and "out-transitions". An "in-transition" starts when the signer begins to create the manual signal for the target word, and ends when they have finished forming the relevant manual signal. An "out-transition" starts when the signer releases the manual signal for the target word, and ends the moment before they form the manual signal for the next word. Figure 1 shows the defined categories that make up sign language, with "signs" and "transitions" plotted along a timeline.

 ISBN: 978-1-61208-325-4

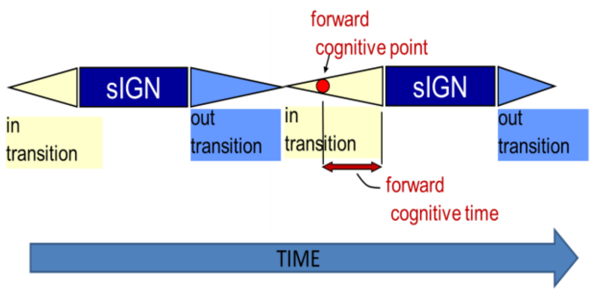


Figure 1. The temporal structure of sign language

III. COGNITIVE EXPERIMENT ON SIGN LANGUAGE WORDS

A. *Compiling experiment data*

We compiled the experiment data (sentences) with the aim of ascertaining the point at which the subjects recognized the target sign language word. Expressions in sign language have various different characteristics. The chosen method can therefore have a significant effect on the results of cognitive experiments. We carried out a preliminary experiment using selected sentences from KOSIGN Ver.2, a dialog corpus that includes 10 sign language words. As the sentences were parts of dialog however, their contents depended on questions asked by another person, which influenced reading of the sign language. Being dialog also meant that the sign language included elements such as agreement, denial and acknowledgement. This created further issues because it changed the position of the target word within the sentences. With that in mind, we began to look into conditions for the compilation of sentences that we could use in the main experiment.

1) *Exploring target words*

We started by determining the position of the target word, as the third word in each sentence. We were unable to position the target word at the start of the sentence because it would have included transitional movements from the initial position, making it impossible to ascertain the point during the out-transition or in-transition at which the subject recognized the word. In natural sign language, the choice of first words is also limited, because the first word in a sentence is often a referent word such as {I} or {you}, or a word indicating the tense of the sentence such as {today} or {yesterday}. The second word in a sentence is easily influenced by the first word, so that would also have limited the range of words available. Taking all of these factors into account, we decided to make the third word the target word, and to use sentences consisting of three or four words.

Next, we selected the target words. Many words in sign language are made up of two or more morphemes. The Japanese word {kazoku} (family) for example consists of the two morphemes {ie} (house) and {hitobito} (people). The in-transition for a word like this is likely to suggest a number of possible words, including {ie} (house) {yane} (roof) or {kazoku} (family). We therefore decided to use words consisting of a single morpheme as candidates for the experiment. We also decided to include sign language for the selected words, as detailed below.

a) One-handed sign language: {omou} (think)

b) Two-handed sign language: {hana} (flower)

c) Different sign language with each hand: {odororku (surprise)

d) Sign language with changing handshape: {nenrei} (age)

e) Sign language with classifier handshape: {asa} (morning), {neko} (cat)

To enable the experiment to be carried out on both native signers and on non-native signers, we selected words that even a beginner just starting to learn sign language would know.

2) *Compiling experiment sentences*

The spatial structure of sign language consists of elements such as handshapes and articulatory gestures. To understand the way in which subjects recognize sign language words in practice, we looked at three different types of example sentences; Type A, Type B and Type C.

We selected the target words for Type A sentences so that the sentences would be grammatically and semantically correct in the context of sign language. We called these "correct sentences". The following is an example of the correct sentences used in the experiment. The words between curved brackets { } are sign language words. The words between speech marks " " give the meaning of the sentence in English.

Example correct sentence:
{fujisan} {miru} {asa} {kirei}
(Mount Fuji) (look) (morning) (beautiful)

English translation:
"Mount Fuji looks beautiful in the morning"

Type B sentences had the correct grammatical structure, but we selected target words that would affect how the sentence was interpreted. We called these "difficult to interpret sentences". For Type B sentences, we tried to select target words whose sign language movements were dissimilar to the target words used in the Type A sentences.

Example difficult to interpret sentence:
{ashita} {tenki} {ame}
(tomorrow) (weather) (candy)

English translation:
"Tomorrow the weather will be candy"

Type C sentences contained target words consisting of fabricated movements based on phonemes that are absent from Japanese sign language ("pseudowords"), making it impossible for the reader to interpret their meaning. We called these "pseudoword sentences". We complied the sign language movements for Type C sentences so that the transition from the second word to the target word resembled the equivalent transition in the Type A sentences.

Example pseudoword sentence:
{watashi} {ototo} {pseudoword 1}
(I) (younger brother) (pseudoword 1)

In this sentence, the sign language movement for "pseudoword 1" starts by moving from the cheek, as in the sign language word {shumi} (hobby), before opening and lowering the hand, as in the sign language word {hikari} (light). The aim of using a pseudoword is to check that the subjects are taking the reading experiment seriously.

3) Filming clips of example sentences

We compiled 226 correct sentences, 48 difficult to interpret sentences and 12 pseudo word sentences for use in the reading experiment. We then filmed a native signer signing the sentences. For the purpose of the reading experiment itself, we selected 20 correct sentences, two difficult to interpret sentences and one pseudo word sentence from the sign language sentences we had recorded. Table 1 shows the selected sentences.

We used the following procedure to edit the sign language sentences for the reading experiment.

a) We took the section from the start of the sentence to the first frame of the out-transition from the second word, immediately before the target word, and called it "Pattern 1" (Figure 2(a)). All Pattern 1 clips cut off at the out-transition from the second word.

b) Next, we added the next frame following on from Pattern 1 and called the resulting clip "Pattern 2" (Figure 2(b)). So Pattern 2 clips kept running for one frame longer than Pattern 1 clips.

c) We continued to produce similar patterns, by adding another frame at the out-transition from the second word, one frame at a time.

We continued this process of adding one frame at a time, all the way through to the out-transition from the target word. The final experiment clip was Pattern n, which showed the target word in its entirety.

Figure 2 shows the progression through the different patterns used for the experiment clips. If subjects had been given a long time to think, they might have worked out the end of the sentence. We therefore edited the clips so that the patterns played through continuously, one after another. As subjects required some thinking time however, we edited the clips with a three-second countdown between each pattern.

B. Experiment procedure

We explained to the subjects in advance that we wanted them to recognize the third word in each sentence. To prevent subjects from working out the answers from the context, and from being influenced by the Japanese language, we asked them to watch each clip and then give their answers in sign language during the countdown before the next clip. We then recorded the experiment to determine at what point (pattern) the subjects were able to read the target word. If a subject was unsure about the target word or gave an incorrect answer, we continued to play the clips until they gave the correct answer. We recorded subjects' answers even if they were incorrect. Correct, difficult to interpret and pseudoword sentences were shown in a random order, rather than the order listed in TABLE I (next page).

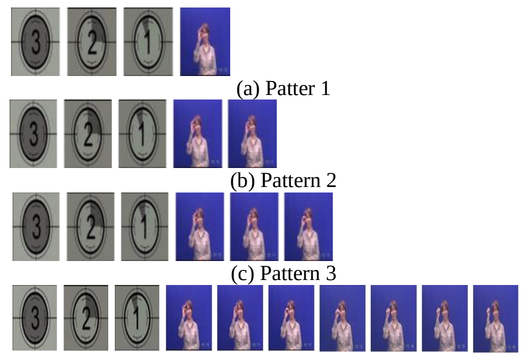


Figure 2. Example patterns for the experiment sentences

C. Subjects

A total of 13 native-signers took part in the experiment as subjects, as outlined in TABLE II. It is difficult for to the cooperation of the experiment for a hearing impaired people. In addition, this experiment takes long time. Therefore I have become 13 people in this article.

TABLE II. LIST OF SUBJECTS

Subject	Age group	Gender	Age when Subject Lost Hearing
A	60s	Female	0
B	30s	Female	4
C	20s	Male	3
D	50s	Female	5
E	60s	Female	0
F	40s	Male	3
G	40s	Female	0
H	60s	Male	1
I	60s	Female	0
J	20s	Male	0
K	40s	Female	0
L	50s	Female	0
M	50s	Female	0

TABLE I. LIST OF SENTENCES USED IN THE EXPERIMENT

Sign language words		1st	2nd	3rd	4th
Correct sentences	1	fujisan (Mount Fuji)	miru (look)	asa (morning)	kirei (beautiful)
	2	kore (this)	yoi (good)	omou (think)	
	3	isshokenmei (to the best of one's ability)	sagasu (search)	shikashi (however)	nakatta (not there)
	4	taberu (eat)	ato (after)	tsugi (next)	nani (what)
	5	ima (now)	kyuryo (salary)	sukunai (low)	
	6	kino (yesterday)	neko (cat)	shinu (die)	kanashii (sad)
	7	ima (now)	nani (what)	tsukuru (make)	anata? (you)
	8	moshikomi (application)	ashita (tomorrow)	made (by)	daijyobu? (OK)
	9	watashi (I)	okane (money)	nai (not have)	watashi (I)
	10	kyo (today)	miru (watch)	dake (only)	owaru (end)
	11	anata (you)	kao (face)	odoroku (surprise)	nani? (what)
	12	massugu (straight ahead)	iku (go)	mura (village)	aru (there is)
	13	hon (book)	kasu (lend)	kamawanai (OK)	
	14	watashi (I)	suiei (swimming)	tokui (good at)	watashi (I)
	15	anata (you)	chichi (father)	toshi (age)	ikutsu? (how old)
	16	senshu (last week)	doyobi (Saturday)	shibai (play)	mita (see)
	17	anata (you)	suki (like)	hana (flower)	nani? (what)
	18	sono (that)	heya (room)	tabako (cigarette)	dame (forbidden)
	19	kino (yesterday)	yoru (night)	msuume (daughter)	kaetta (returned)
	20	anata (you)	shuwa (sign language)	dekiru (can do)	
Semantically unconventional sentences	1	ashita (tomorrow)	tenki (weather)	ame (candy)	
	2	watashi (I)	musuko (son)	neko (cat)	kaicho (chairman)
Phonologically unconventional sentences	1	watashi (I)	ototo (younger brothre)	shumi(opening hand from fist as in {hikari}(light) (hobby)	gemu (video games)

IV. COGNITIVE EXPERIMENT ON SIGN LANGUAGE WORDS

Table III shows the point at which the 13 subjects recognized the target word in each of the sign language sentences. The position of the out-transition and in-transition stages was determined by the native signer who compiled the example sentences.

Table III shows that, when faced with correct sentences, subjects answered most target words correctly between the out-transition from the second word and the in-transition to the third word.

Figure 3 shows the cumulative rate of recognition of target words at three points in time; during the in-transition, out-transition and word movement.

As Figure 3 clearly shows, most of the subjects recognized the target word before the target word movement. The rate of recognition also increased dramatically during the first 80% (approx.) of the time following the start of the out-transition from the second word. During the first 30% (approx.) of the time following the start of the in-transition to the target word meanwhile, the rate of recognition was over 50%.

TABLE III. RECOGNITION TIMES FOR TARGET WORDS

Example sentences	Target word	Result (recognition)		
		Out-transition from second word	In-transition to third word	Movement for third word
Correct sentences	asa (morning)	0	1	12
	omou (think)	13	0	0
	shikashi (however)	1	7	4
	tsugi (next)	1	9	2
	sukunai (low)	1	12	1
	shinu (die)	7	6	0
	tsukuru (make)	3	9	1
	made (by)	13	0	0
	nai (not have)	13	0	0
	dake (only)	13	0	0
	bikkuri (surprise)	0	9	4
	mura (village)	0	13	0
	kamawanai (OK)	13	0	0
	tokui (good at)	8	5	0
	toshi (age)	7	6	0
	shibai (play)	8	5	0
	hana (flower)	0	5	8
	tabako (cigarette)	7	6	0
	musume (daughter)	0	9	4
	dekiru (can do)	13	0	0
Semantically unconventional sentences	ame (candy)	0	10	3
	neko (cat)	1	10	2
Phonologically unconventional sentences	Non-sign language (shumi (hobby))	0	4	9

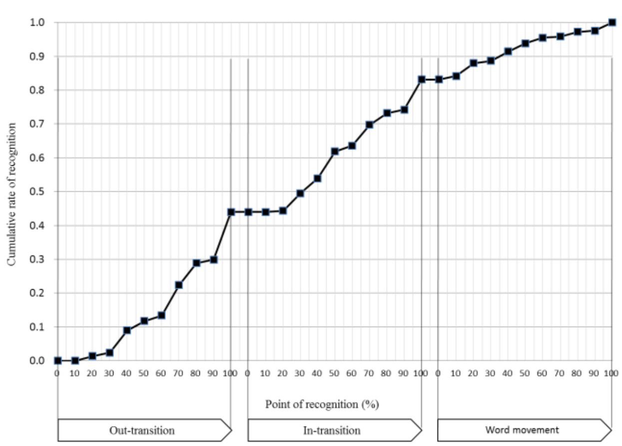


Figure 3. Cumulative rate of recognition in correct sentences

We also ran the experiment using difficult to interpret and pseudoword sentences, to examine how subjects were able to recognize words during the transition stages. In the case of difficult to interpret sentences, 3% of subjects recognized the target word during the out-transition, 67% during the in-transition, and 30% during the word movement, indicating that 70% of subjects still recognized the target word during the transition stages. The target words in the difficult to interpret sentences were sign language words whose meaning was unpredictable. It is therefore to be expected that recognition amongst the subjects would be slightly slower than normal. Although there were few correct answers during the out-transition from the second word in the case of difficult to interpret sentences, the rate of recognition shot up to 67% during the in-transition to the target word. This suggests that native signers are able to recognize words based on the transition, even when dealing with difficult to interpret sentences.

If we look at incorrect answers given in response to correct sentences, we see that the subjects incorrectly recognized very similar words as the target. The following section examines two typical examples in sign language.

Example incorrect answer 1:

{asa} → {mitomeru} {mata}
(morning) → (recognize) (again)

In the first example, the movement for the in-transition is similar. This is probably why many of the subjects recognized the target word during the word movement for the third word.

Example incorrect answer 2:
{sukunai} → {yasui} {mata} {binbo}
(few/low) → (cheap) (again) (poor)

The second example contains a different movement for the in-transition. The sign language for the target word was supposed to be the following.

Example incorrect answer 3:
{ima} {kyuryo} {sukunai}
(now) (salary) (low)
English translation:
"My current salary is low"

It is likely that the subjects inferred from the second word {kyuryo} (salary) that the next word would be {yasui} (cheap) or {binbo} (poor).

At the same time, there were some words that more or less all of the subjects recognized during the out-transition from the second word. The words in question – {made} (by), {dake} (only), {nai} (not have), {kamawanai} (OK) and {dekiru} (can do) – all have characteristic word movements that are easy to recognize.

Based on these results, we can assume that native signers predict the next word in a sentence partway through the transition stages, based on information such as expressions and intonation.

V. CONCLUSION AND CHALLENGES FOR THE FUTURE

This paper is an experimental study into the time taken for deaf people to predict and understand words at the end of sentences when reading Japanese sign language.

We divided the sign language sentences used in the reading experiment into three different types depending on the nature of the word at the end of each sentence (the "target word"). Type A sentences ("correct sentences") contained target words that enabled the reader to correctly interpret the meaning of the sentence. Type B sentences ("difficult to interpret sentences") contained target words that were of the correct type syntactically, but that did not enable the reader to correctly interpret the meaning of the sentence. Type C sentences ("pseudoword sentences") contained target words consisting of fabricated movements based on phonemes that are absent from Japanese sign language ("pseudowords"), making it impossible for the reader to interpret their meaning.

For the experiment itself, we asked native signers to read sign language sentences following three different patterns. The results showed that, in the case of correct sentences, over 80% of the subjects predicted and correctly understood the target word before the word itself was signed. In the case of difficult to interpret sentences meanwhile, 70% of subjects predicted the target word before the word itself was signed and were able to interpret the sentences in spite of their unconventional meanings. When dealing with pseudoword sentences however, the subjects were unable to predict the signer's movements until the fabricated movement for the target word was performed.

Based on these results, we can assume that deaf people predict words and understand sentences by effectively utilizing the information required to form correct sentences, including expressions and intonation.

The challenge for the future is whether this method of reading information can be learnt through sign language education. With the clarification of ascertaining the point at which the subjects recognized the target sign language word, that can be helpful in generating sign language animation.

ACKNOWLEDGMENT

We would like to thank everyone who assisted with the recording of the sign language clips and who took part in the experiment.

REFERENCES

[1] Ministry of Health, Labour and Welfare of Japan:"Survey on persons with pyhsical disability",2006(in Japanese).
[2] C. Baker:" Regulators and turn-taking in American Sign Language discourse", In:L.A.Fridman(Ed.), On the other hand, New York: Academic Press, pp.215-236,1977.
[3] K.Nakazono, et.al :" Quantification Method of Delay Effect on Sign Dialogue", The IEICE Transactions on Information and Systems, Vol.J90-D, No.3, pp. 1-11, 2007(in Japanese).
[4] M.Terauchi, K.Nakazono and Y.Nagashima:" Fundamental Analysis of Sign Recognition during JSL Sentence Reading", IEICE technical report, Vol.107, No.369(HIP2007-145), pp. 91-95, 2007(in Japanese).

Aurora - Exploring Social Online Learning Tools Through Design

Peter Purgathofer and Naemi Luckner
Vienna University of Technology
Institute of Design and Assessment of Technology
{purg, naemi}@igw.tuwien.ac.at

***Abstract*—Teaching is an integral part of our work at university and we strive to achieve a high quality in teaching our students. Yet, we face classes with up to 800 students per semester and do not have the resources to ensure a close rapport with each and every one of them. For the last 6 years we have been working on an online solution for this problem, with the aim of letting students take responsibility of a large part of their own learning process. The e-learning system Aurora envelops a number of components ranging from organizational and informational tools to discussion systems and student portfolios. Students are invited to participate actively in this online learning environment and are presented with a variety of options they can choose from in order to accumulate enough credits to pass their courses. Over the years of developing Aurora, issues and inadequacies of the system became apparent and lead us to change our system design iteratively, learning from its shortcomings.**

***Keywords*-Asynchronous Interaction, E-Learning, E-Portfolio, Electronic Note Taking, Backchannel, Teaching, Design, Newsfeed**

I. Introduction

The Web 2.0 as described in [1] changed online culture and shifted it from a passive consumer culture to a participatory culture. This development also influenced the process of teaching and learning, which is since referred to as E-Learning 2.0. The notion of E-Learning 2.0 is that Web 2.0 technologies are adapted and integrated in E-Learning systems [2]. Knowledge can be created, shared, remixed and repurposed by communities of practice. Students are part of this process, collect sources and participate in the communities, by sharing their own ideas and findings. Brown and Adler [3] describe a new age of education, in which lifelong learning is not only needed but also supported by the participatory architecture of the Web 2.0. They speak of a new learning approach, "...characterized by a demand-pull rather than the traditional supply-push mode..." of obtaining knowledge. They emphasize the importance of social learning in the new online learning environment, pointing out, that the traditional teacher-student relationship is exchanged by a peer-based learning relationship.

Siemens [4] took this change of learning culture and devised his theory of connectivism. He states that learning in the digital age is the self-driven process of building up networks of knowledge. Nodes in a network can be data sources, communities or people and are connected to the network with strong and weak links. Weak links are more interesting since they can open doors to new areas of knowledge, diversity and innovation. Siemens points out that the lifecycle of 'correct' facts is getting shorter, and new knowledge is created faster, so memorizing facts is not yielding desired results anymore. More important is the 'Know-where', which describes where knowledge can be found quickly rather than learning the knowledge itself.

In this paper, we present an E-Learning System with the aim of letting students take responsibility of their own learning process. The system is an attempt to create a holistic learning platform, valuing not only assigned course work, but also social and other additional content students create or discover over the course of a semester. What we wanted to avoid was to develop another system increasing the distance between teacher and students. Instead, our goal was to start from the rather difficult situation of very large classes, where contact between teacher and student is short and far between, and transform it so that students have a feeling of more immediate involvement, more contact and more personal mentoring. To achieve this, we put concepts like social interaction, participation, and exchange at the center of our design efforts.

Aurora is a learning platform consisting of three modules that can communicate with each other. The Dashboard is an administrative tool, containing an administrative Newsfeed as well as widgets to enhance communication between all participants of the course and maintain an overview of the course progress and interesting developments around the course topics. The Slides module is used during and after lectures as backchannel and basis for upcoming discussions around course topics. Students are provided with a pool of activities they can choose from in the Portfolio. We chose the word activity rather than exercise for work assignments, since we want to motivate students to actively pursue their work for this course and we wanted to avoid the vocabulary usually associated with course work to try to increase motivation. The name Aurora is not an acronym, nor has it any deeper meaning. We used the name because it refers to something beautiful, and because it sounds good. Also, pictures of auroras give awesome visuals to be used in the layout.

The remainder of this paper is structured as follows: The next section 'Overall Goal' will give an idea of our motivation and process to create a new e-learning system and explain why we chose to develop the system ourselves instead of taking existing tools. Subsequently, in the section 'Components', each module of Aurora is described and compared with existing solutions from literature. The description is followed by a preliminary evaluation, 'Evaluation', and finally a conclusive section, 'Conclusion and Future plans', in which we describe how the evaluated data influences future designs.

 ISBN: 978-1-61208-325-4

II. Overall Goal

At the Vienna University of Technology, lecture participants are sometimes in the high three-digit numbers. Traditionally, this would mean that lectures have to be endured with a passive and consuming stance. Around 2005, we set out to explore new ways to make lectures more interactive. We started by appropriating existing systems like IRC and Twitter to facilitate backchannel communication and interaction for students visiting large lectures. Early on, we were fascinated by the idea that we could time-sync this information to the slides. This would enable students to understand the backchannel as a means of taking (collaborative) course notes that became attached to the individual slides of the lecture.

We also started to replace the then prevailing passive html web pages for course information with blogs, which seemed an ideal fit for some years. Later, we started to supplant blogs with a custom-made newsfeed solution that was heavily inspired by the structures and aesthetics of social media systems like Facebook or Twitter.

By replacing all the passive elements of large scale courses with interactive components, we also set out to change the way we evaluate student performance in order to come to a final grade. This led to a somewhat idiosyncratic redefinition of a portfolio system that we implemented.

All these systems are currently being actively developed and refined in an effort to explore new ways of teaching and learning for a generation that grew up with ever-present internet access and for the most part played a lot of games [5]. We redesign our systems year after year after understanding what works and what doesn't. We pursuit this research in the spirit of design as research, or explorative design. One core idea is that with each version, new concepts become evident that were impossible to see last year, be it from use, from formal or informal evaluation, or because we reflect on our progress from the feedback we get from students.

Following this path for some years now, we have come to a place where individual components have been published about, but we never set out to describe the system as a whole. This is what this paper sets out to do.

III. Components

A. Dashboard

Dashboards are often used in complex system to provide participants with an overview of activity on these platforms. The role of a dashboard is variable, depending on the context of its application. Dashboards have been used to track activity from different applications in a complex system [6]; to create peripheral awareness, provide navigation and a system-wide inbox [7]; to create awareness of group members' actions and to convey the status of shared artifacts [8]; and to provide multiple views of a large dataset in a system [9]. More specifically, in an e-learning context, they have been used for self-monitoring for students and to improve teachers' awareness [10]; and to help students to relate their learning experience to that of their peers or other actors in the system [11].

In Aurora, the Dashboard is the first page every student is presented with when logging into the system. It is a collection of widgets, containing the Newsfeed, an individual course status overview, showing colleagues, groups, current links and additional contact information. The page draws together course-relevant information related to the content from other websites, as well as information from other components of Aurora.

In former versions of Aurora, we included a statistics page to enhance students' peripheral awareness. The page provided a statistical overview of the large amount of data that is distributed over the whole system. Students could for example look up who of their peers was involved in a lot of discussions, or who got a lot of stars, which could be awarded for good comments by other students and members of the staff. This view has not made it into current versions of the system, mostly because of a lack of time and resources for the development.

1) Newsfeed: The Newsfeed is a largely organizational message board, but can also be used for content related postings. The lecture staff can use the Newsfeed to publish course updates and other relevant news for the students. Questions, annotations, complaints and praise can also be posted here, and can be answered by other actors in the system. Students post content related comments as well, but are asked to first look for a suitable slide in the Slides section to provide context for the content, before blindly posting it in the Newsfeed.

Information from other components is collected and posted via sticky notes at the top of the Newsfeed. Students are informed if someone answered to one of their postings in the Slides section and can jump directly to the posting via link. If students get points for a good comment in the Slides section or for a newly marked activity in the Portfolio, they are notified here. Direct messages show up on top of the Newsfeed section, and can be sent by either colleagues or team members.

The Newsfeed enhances direct communication between students and staff and also provides a forum for discussions about the course design. It can be searched or filtered to see either only staff postings, only organizational postings, or only content related postings. Students can subscribe to Newsfeed postings via RSS to integrate them into their everyday online environment.

2) Additional widgets: The Progress Bar widget is a tool students can use to get an overview of their progress in each of their classes. Each lecture has an overview of the student's activity status. It shows the amount of points received in the lecture through activities and through comments, as well as the total amount of points. Additionally, it shows the amount of points that the student has handed in but which have not yet been graded and the amount of points the student can still hand in until the end of the semester.

In the Colleagues widget, users can add other students to their course network and, on accept, see their avatars and further information. They can write direct messages to their colleagues as well as see all their colleagues' comments in the Newsfeed and the Slides highlighted. This can create a feeling of connectedness within the course and motivate to interact with others regularly.

Some activities in the Portfolio can be worked on in teams. The Teams widget shows a list of all existing teams the student is a part off. Each entry contains the name of the project the team is working on, the possibility to send a message to all team members, and a list of the other team members.

The Current Links widget displays a list of recent articles and interesting websites - supplementary reading material of topics covered in the course. The collection of links is compiled in a blog using soup.io and integrated into the Dashboard via RSS.

Lastly, the Dashboard lists contact information to correspond with the staff directly. Students are invited to ask all course relevant questions directly in the Newsfeed so that other students can profit from the answers as well, but some issues need to be taken up with the staff directly.

B. Slides

There is some research on how to offer interactivity in large lectures. One approach are the Audience Response Systems, also called Clickers. Kumar and Rogers pioneered such systems in their 1976 Olin Experimental Classroom [12] that featured a feedback channel for students in the form of 12 buttons. Today, clickers are commercially developed products, offering a number of potential benefits to large lectures. Caldwell [13] summarized the literature on using clickers in lectures. Recently, software clickers based on the fact that most students bring a network-connected device, most prominently mobile phones, to lectures have begun to appear, but this approach is still mostly experimental [14] [15].

Of course, more elaborate backchannel communication systems have been tried as well, such as ActiveClass [16], Fragmented Social Mirror [17] or ClassCommons [18]. The development and evaluation of these systems overlaps with the development of the approaches presented here, first published in 2008 [19].

It can be argued that backchannel communication during lectures is potentially distracting, diverting the attention from the speaker to unrelated things. On the other hand, students regularly bring their laptops to class in the hope of finding productive use, but often end up doing other stuff that is available on the computer. We have observed that supplying students with a backchannel that is centered around the lecture itself brings some of that attention back, and while it creates bubbles of divertion from the lecture itself, at least these bubbles are focussed on the content of the lecture.

Slides consist of two major components, Livecasting and Studio. Livecasting lets participants add notes to individual slides of a lecture, either in the style of a backchannel conversation, or privately. Once the lecture is finished, slides and comments are available in a combined view in the Studio. Participants can keep adding comments, links etc. in the Studio, so that the lecture slides become the focal point of discussion and exchange for participants and lecturers alike.

1) The Livecasting component (Figure 1): During a lecture, the lecturer runs a script on her computer. By pressing the next slide button on the remote, she triggers a script that sends the number and title of the newly displayed slide to the slidecasting server. Additionally, the script retrieves the lecture notes of this slide in the presentation document and scans them for a custom-made meta-syntax signifying information that is meant to be posted with the same slide. These text-lines include explanations, enhanced quotes, references and other links, activities and discussion starters.

Participants load a web page that changes with each slide the lecturer shows, so that information entered into either the public comment or private note fields on this page ends up being attached to the slide that was visible when the participant started typing. All participants can see the public comments entered by other participants, and they can reply to these comments, creating ad-hoc discussions of the lecture content. To ease the cognitive load, a participant's own comments are colored yellow. Additionally, students have the opportunity to mark slides as liked, important or unclear with a single click.

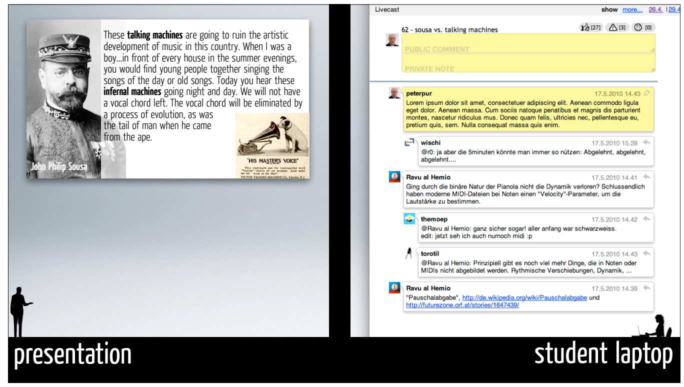


Figure 1: Livecasting setup of the Slidecasting system

2) The Slidecasting Studio (Figure 2): Once the lecture is over, the lecturer makes slide-by-slide pictures of his presentation available to the slidecasting system. While this can also be done before the lecture, we decided not to show the slides because of the obvious spikes in network traffic this would generate whenever a new slide is shown. The slides, all the participants' comments as well as the lecturers automatically posted comments are then made available in the Studio.

Here, participants and lecturers can post comments even after the lecture is finished. In the Studio, the slides are arranged horizontally, sorted by their time of appearance in the lecture. The comments attached to each slide are laid out vertically, with the earliest comments up on top (usually the comments posted by the script on the lecturers computer), with reply threads sorted in the same way.

Participants can give praise to good comments by clicking the star next to the avatar of the author, in which case the star turns yellow and shows the number of clicks it has accumulated. Lecturer can use this same mechanism to award points to outstanding comments. In this case, the star is distinguished with a green glowing outline, making its commendation visible to everybody.

While lecturer's comments are generally displayed in the same way as student comments, there are two lecturer-posted types that stand out from the rest: discussion starters and activities. Comments of both these types are arranged between the slide and the private notes delineator, thus standing out even when scrolling through the slides quickly.

Discussion starter comments typically contain a questions and an invitation to discuss this question in the comments of the slide. We use this mechanism to post slides in between lectures, asking participants to discuss upcoming content. Activities contain a brief explanation of an activity, linking into the Portfolio system where an elaborate description of this activity can be found. This gives the lecturer an opportunity to announce new activities that derive from the content on a slide.

Activities take student to the Portfolio of Aurora, where they hand in their work for review and evaluation.

C. Portfolio

In areas like HCI or Informatics and society, it is hard to make written exams, and once you have more than a couple of hundred students, it becomes impractical to the point of impossible to conduct oral exams. We started to abandon tests

 ISBN: 978-1-61208-325-4

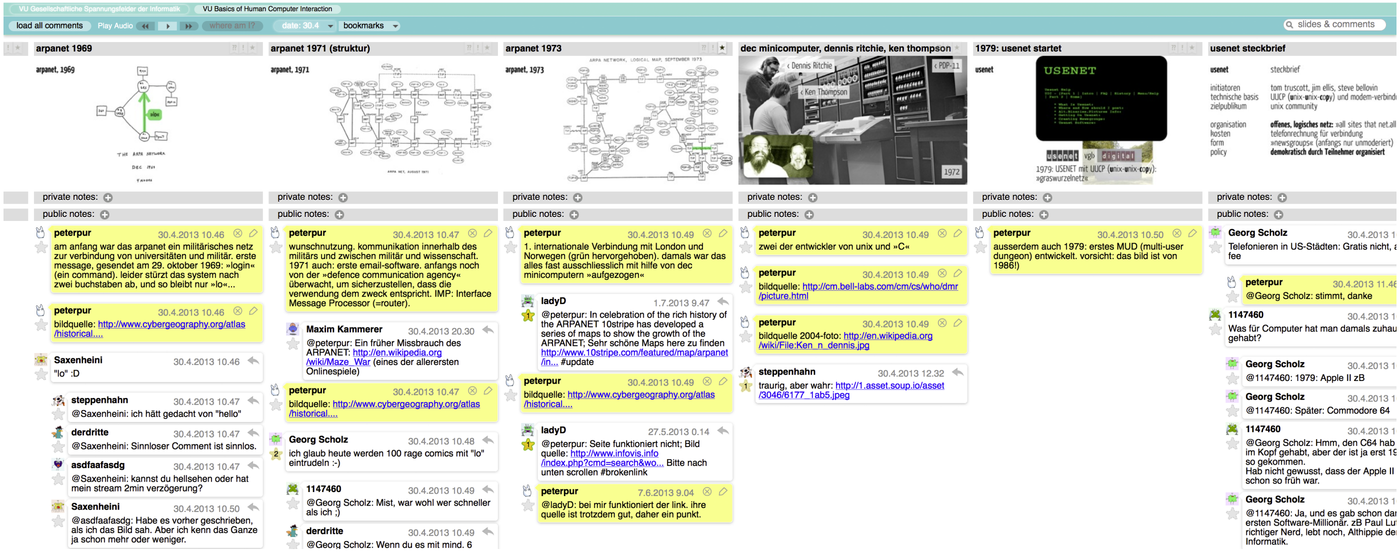


Figure 2: Slidecasting Studio, where all slides and comments become accessible to participants and lecturers alike

and exams at some point when we had the distinct impression that the fact that we had to make a written exam changed what we taught. This compromises the whole idea of teaching, especially at the university level.

For a couple of years now, research papers have been explaining the theoretical sense the adoption of ePortfolios would make. Advantages implied are, among others, improved reflection, increased student engagement, improved learning outcomes, and increased integration of knowledge [20]. The paper quoted gives a comprehensive overview over ePortfolio research, and points out the lack of empirical support for many of the asserted advantages.

The module we call Portfolio is not really an ePortfolio in the strict sense of the word. While we explicitly ask the students put upload artefacts that shows what they have learned, we offer a large catalog of predefined activities that can be handed in here. These activities include a broad range of tasks, from simple applications of theoretical content, to actions reflecting their own prior projects, to complex design exercises. Many of those activities would make viable exercises in a traditional deadline-based context, while others would be quite unsuitable for such an environment. The catalog also contains meta-activities such as finding new sources, suggesting new activities, and organizing round table discussions with experts in the field. No activity should yield substantially more than 10% of the final grade, so that students will be exposed to a broad range of topics.

Participants hand in their work using the portfolio system of Aurora. We do not set any deadlines other than the end of the semester, and we do not expect them to follow a specific order. The only requirement they have to meet is to make sure that their work is distributed throughout the semester, instead of congested at the end.

The Portfolio includes an easy review component for the course admins to review and evaluate the participant's work, with the notable addition of enabling for the students repeated submission of work that failed to meet the standards. It also includes a double blind peer review component that makes part of the assessment process into an activity by itself on the premise that if you do an honest review of somebody else's work, you will learn a lot.

This approach tries to abandon the usual scattering of deadlines through the semester, giving the students a lot of autonomy in their work, which self-determination theory deems essential for intrinsic motivation [21].

IV. Evaluation

Our focus in evaluating these components is in better understanding how we can advance the system. We do not have an ultimate goal, but we use both the design process and the evaluation to understand how the system should be enhanced, refined and changed in order to satisfy our needs as teachers as well as the needs of the students as learners.

TABLE I: The table shows how many people were involved in the evaluated courses, staff as well as students. Additionally, it shows how many certificates were handed out for each course at the end of the semester.

	Profs	Predoc	Tutors	Students	Certificates
BHCI	3	1	6	733	442
IST	1	1	4	521	337
Total	3	1	10	1254	779

This preliminary evaluation is based on data from two courses, Basics of Human Computer Interaction (BHCI) and Interactions of Society and Technology (IST), which took place in the summer semester of 2013. A total of 11.793 activities was handed in over the course of the semester, 7126 in BHCI and 4667 in IST. The staff of both courses combined consisted of 3 professors, 1 predoctoral fellow and 10 tutors, exact numbers can be found in Table I. Students only got a certificate if they handed in at least one activity. Every student who ultimately received a certificate handed in 15 activities on average. In the Slides section, 1283 slides were posted distributed over two courses with 23 lectures in total, and 3975 comments were written during and after these lectures.

 ISBN: 978-1-61208-325-4

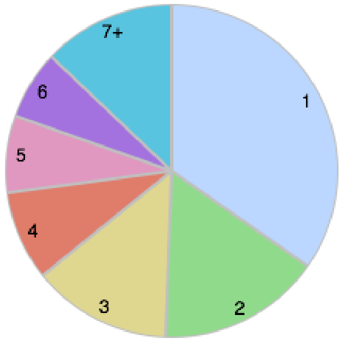


Figure 3: Time it took to grade an exercise, calculated in weeks

A. Portfolio evaluation

Figure 3 shows a pie chart of the time it took to grade activities. One third of the activities were graded after a week, which would be an acceptable amount of time for students to wait for feedback. Given the student-staff ratio, we tried to achieve a maximum waiting time of three weeks until every activity is graded. As can be seen in Figure 3, we were not able to reach that goal, as only two thirds of handed in work was evaluated within the given time frame. The final third of the pie chart consists of activities that took 4 and more weeks to be graded. Considering the importance of feedback in order to keep students motivated and continuously working [22], 4+ weeks seems too long a time to hear back on one's work.

We suspect that this fluctuation in delay can actually be explained by queue modeling in game theory. Activities tend to be handed in unequally distributed in time, leading to an overload what causes congestions that are then almost impossible to resolve within the given resources until the end of the semester.

B. Slides evaluation

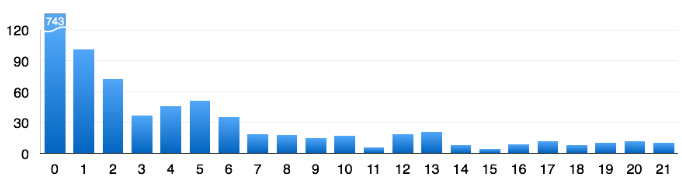


Figure 4: How long after a slide was posted (Day 0) are students interacting with it via the comment stream.

In Figure 4, the comment data was analyzed to find out if and how long after the lecture students engage with the content by writing comments and discussing it in the Studio. The graph shows that most comments are written during the lecture, but there is a long tail (going up to 75 days) after the original slide was presented. Approximately 120 comments were posted even after the semester was over.

Also interesting for us is the second peak a couple of days later, as well as the 'long tail' of posted content after the lecture that can be seen in Figure 4. A cursory evaluation of these 'late' postings show that students come back to post things they find relevant, like news coverage, examples, references, etc. or to partake in discussions they have started with another postings. We see this as a successful feature of the system, as it induces reflection of and occupation with the content of a lecture quite some time after the lecture is over.

V. Conclusion and Future plans

The main goal of our work is to explore the design space of online teaching and learning support systems. Our approach is best described as explorative design, with the main goal to better understand the situation, the players, and their needs. In building and using systems that implement novel approaches to the context of teaching and learning, we in turn have a chance to understand the change such systems bring into the situation, and react accordingly. This approach shifts the focus of evaluation from understanding how and why the approach worked (or not), to finding new approaches to try. In the end, we are not interested in proving that our approach is right, e.g., by showing effectiveness by some abstract learning measurements. Instead, we want to find new and better ways to teach and learn that use the potentials of new technologies, and tap student's self-motivational capabilities.

The next version of Aurora is already in development. We redesigned Portfolio in order to address previously identified problems, and to explore new ways of organizing evaluation and grading. Our main goal was to reduce the students' waiting time for graded activities by introducing double blind peer reviews into the grading process. Also, activities will be organized by topic and will be structured in 'levels' that build on each other. Only top level activities are considered grade relevant and will thus be graded by a staff member. All other activities will just be reviewed by two or more peers with an emphasis on constructive feedback, not marks, given by the reviewers. For every activity a student hands in, they have to review three activities of the same type handed in by colleagues. To maintain a certain level of quality, activities as well as reviews will randomly be checked by members of the staff. We will also implement an easy way for students to report extremely poor elaborations, plagiarized work, or, on the other side, if they did not get meaningful feedback from a reviewer.

The second component that is undergoing major changes is the Newsfeed. We are putting more emphasis on ways to filter and reorder the comment stream in order to make information better accessible. This redesign is based on the way reddit [23] works, e.g., using up- and downvoting to incentivize collaborative content filtering. We also designed a feature to extract good (student) questions and corresponding (student or staff) answers from the Newsfeed for display in a separate FAQ section.

Another step in Aurora's development will be to include our discussion component Discourse into the system. Discourse is a threaded asynchronous discussion forum that differs from traditional systems in the way discussions are represented. It vertically displays new thoughts, arguments and ideas in the discussion, and horizontally the answers and exchanges to said arguments. Students can gain points towards their final grade when participating in discussions and composing clear arguments supported by references.

Handling more than 500 students in university courses is a rare situation. Often, such a challenge is tackled by introducing distance between teachers and learners, and by relying on examination and tests. This removes autonomy from the learning process, which we see as a central property. Thus, we tried to go the opposite way, and designed Aurora with the explicit goal to give students as much autonomy as possible in such a setting. In our experience, such a challenge requires explorative approaches, learning not only from evaluation, but also from the design process itself.

ACKNOWLEDGMENT

The authors would like to thank all contributors who have been involved in the development of Aurora over the years: Stephan Bauer, Elisabeth Bauernhofer, Christoph Brner, Daniel Domberger, Michael Emhofer, Andreas Fermitsch, Martin Flucka, Thomas Gradisnik, Peter Holzkorn, Daniel Kececi, Lucia Leitner, Peter Minarik, Wilfried Reinthaler, Gerald Reitschmied, Diane Salter, Reinhard Seiler, Martin Sereinig, Raif Tabucic, Bruno Tunjic, Wolfgang Zalesak

REFERENCES

[1] T. O'Reilly, "What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software," Design, vol. 65, 2007, pp. 17–37.

[2] S. Downes, "Feature: E-learning 2.0," Elearn magazine, vol. 2005, no. 10, Oct. 2005, pp. 1—-, [Accessed Jan. 24, 2014]. [Online]. Available: http://doi.acm.org/10.1145/1104966.1104968

[3] J. S. Brown and R. P. Adler, "Minds on Fire: Open Education, the Long Tail, and Learning 2.0," Educause Review, vol. 43 (1), 2008, pp. 16–32, [Accessed Jan. 24, 2014]. [Online]. Available: http://www.educause.edu/ir/library/pdf/ERM0811.pdf

[4] G. Siemens, "Connectivism: A Learning Theory for the Digital Age," 2004, [Accessed Jan. 24, 2014]. [Online]. Available: http://www.elearnspace.org/Articles/connectivism.htm

[5] M. Irvine, "Survey: 97 Percent Of Children Play Video Games," 2008, [Accessed Jan. 24, 2014]. [Online]. Available: http://www.huffingtonpost.com/2008/09/16/survey-97-percent-of-chil_n_126948.html

[6] J. L. Santos, S. Govaerts, K. Verbert, and E. Duval, "Goal-oriented Visualizations of Activity Tracking: A Case Study with Engineering Students," in Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge, ser. LAK '12. New York, NY, USA: ACM, 2012, pp. 143–152, [Accessed Jan. 24, 2014]. [Online]. Available: http://doi.acm.org/10.1145/2330601.2330639

[7] C. Treude and M.-A. Storey, "Awareness 2.0: Staying Aware of Projects, Developers and Tasks Using Dashboards and Feeds," in Proceedings of the 32Nd ACM/IEEE International Conference on Software Engineering - Volume 1, ser. ICSE '10. New York, NY, USA: ACM, 2010, pp. 365–374, [Accessed Jan. 24, 2014]. [Online]. Available: http://doi.acm.org/10.1145/1806799.1806854

[8] J. T. Biehl, M. Czerwinski, G. Smith, and G. G. Robertson, "FASTDash: A Visual Dashboard for Fostering Awareness in Software Teams," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '07. New York, NY, USA: ACM, 2007, pp. 1313–1322, [Accessed Jan. 24, 2014]. [Online]. Available: http://doi.acm.org/10.1145/1240624.1240823

[9] M. McKeon, "Harnessing the web information ecosystem with wiki-based visualization dashboards." IEEE transactions on visualization and computer graphics, vol. 15, no. 6, 2009, pp. 1081–8, [Accessed Jan. 24, 2014]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19834175

[10] S. Govaerts, K. Verbert, J. Klerkx, and E. Duval, "Visualizing activities for self-reflection and awareness," Advances in Web-Based Learning â ICWL 2010, vol. 6483, 2010, pp. 1–10, [Accessed Jan. 24, 2014]. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-17407-0_10

[11] E. Duval, "Attention Please!: Learning Analytics for Visualization and Recommendation," in Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ser. LAK '11. New York, NY, USA: ACM, 2011, pp. 9–17, [Accessed Jan. 24, 2014]. [Online]. Available: http://doi.acm.org/10.1145/2090116.2090118

[12] V. K. Kumar and J. L. Rogers, "Student Response Behaviors in an Instrumented Feedback Environment," SIGCUE Outlook, vol. Special, Dec. 1978, pp. 34–54, [Accessed Jan. 24, 2014]. [Online]. Available: http://doi.acm.org/10.1145/1318457.1318461

[13] J. E. Caldwell, "Clickers in the Large Classroom: Current Research and Best-Practice Tips," CBE-Life Sciences Education, vol. 6, no. 1, 2007, pp. 9–20, [Accessed Jan. 24, 2014]. [Online]. Available: http://www.lifescied.org/content/6/1/9.abstract

[14] D. Lindquist, T. Denning, M. Kelly, R. Malani, W. G. Griswold, and B. Simon, "Exploring the Potential of Mobile Phones for Active Learning in the Classroom," SIGCSE Bull., vol. 39, no. 1, Mar. 2007, pp. 384–388, [Accessed Jan. 24, 2014]. [Online]. Available: http://doi.acm.org/10.1145/1227504.1227445

[15] S. Teel, D. Schweitzer, and S. Fulton, "Braingame: A Web-based Student Response System," J. Comput. Sci. Coll., vol. 28, no. 2, Dec. 2012, pp. 40–47, [Accessed Jan. 24, 2014]. [Online]. Available: http://dl.acm.org/citation.cfm?id=2382887.2382895

[16] M. Ratto, R. Shapiro, T. M. Truong, and G. W. Griswold, "The Activeclass Project: Experiments in Encouraging Classroom Participation," in Designing for Change in Networked Learning Environments, 2003, vol. 2, pp. 477–486.

[17] T. Bergstrom and K. Karahalios, "Social Mirrors as Social Signals: Transforming Audio into Graphics," Computer Graphics and Applications, IEEE, vol. 29, no. 5, 2009, pp. 22–32, [Accessed Jan. 24, 2014].

[18] H. Du, M. B. Rosson, and J. M. Carroll, "Augmenting Classroom Participation Through Public Digital Backchannels," in Proceedings of the 17th ACM International Conference on Supporting Group Work, ser. GROUP '12. New York, NY, USA: ACM, 2012, pp. 155–164, [Accessed Jan. 24, 2014]. [Online]. Available: http://doi.acm.org/10.1145/2389176.2389201

[19] W. Purgathofer, Peter Reinthaler, "Exploring the Massive Multiplayer E-Learning Concept," Ed-Media Invited Talk, 2008, pp. 1–9, [Accessed Jan. 24, 2014]. [Online]. Available: https://igw.tuwien.ac.at/designlehren/exploring_for_edmedia.pdf

[20] L. H. Bryant and J. R. Chittum, "ePortfolio Effectiveness: A(n Ill-Fated) Search for Empirical Support," International Journal of ePortfolio, vol. 3, no. 2, 2013, pp. 189–198, [Accessed Jan. 24, 2014]. [Online]. Available: http://www.theijep.com

[21] E. L. Deci and R. M. Ryan, "Motivation, personality, and development within embedded social contexts: an overview of self-determination theory," in The oxford handbook of human motivation, R. M. Ryan, Ed. Oxford, 2012, pp. 85–107.

[22] A. P. Rovai, "A constructivist approach to online college learning," The Internet and Higher Education, vol. 7, no. 2, 2004, pp. 79 – 93, [Accessed Jan. 24, 2014]. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1096751604000144

[23] "reddit.com," [Accessed Jan. 24, 2014]. [Online]. Available: http://www.reddit.com/

Continous Learning Feedback

Shaping Teaching through Realtime Feedback

Christian Grund Sørensen
Center for Computer Mediated Epistemology
Aalborg University
Aalborg, Denmark
cgs@km.dk

***Abstract* — In most situations where teaching is involved, the need for learner feedback is of utmost importance. Rhetorical theory of communication (concepts of aptum and kairos) underscores the significance of understanding foreknowledge (Gadamer). Since foreknowledge is by nature assessed prior to the learning experience and observed learning outcomes are assessed after the experience, a temporal gap in assessment is obvious. The aim of this paper is to suggest a method for observing learner attitudes, interests and level of learning in realtime during the learning experience. This is exemplified through the use of the Conceptual Pond application to facilitate an insight for the educator into the process of learning and immersion of the learner. Two use cases are discussed in detail as well as the consequences for teaching.**

Keywords- education; assessment; Conceptual Pond; kairos; aptum; continous

I. Introduction

Educators at every level of education know that one of the crucial points in achieving sustainable learning is recognizing the *scena* (audience) of the teaching event. Regardless of whether the teaching methods are traditional teaching in class rooms, technology enhanced learning, blended learning, or standalone online learning, the question still applies: Who is the recipient of my transference of knowledge and invocation of reflection? The identification of the potential the learner is a key issue in the research on cognitive styles [1].

The answer to this question seems straightforward as it may be an ideal for every educator to be thus informed. Nevertheless, this initial knowledge is only the tip of the iceberg as there may be growing level of knowledge, as well as a change in interest and reflection during the course of the learning activity. In this way, a temporal component adds to the complexity of knowing and understanding the cognitive position from which the learner is prone to receive learning. The Greek philosopher Heraclitus of Ephesus illuminates this temporal complexity by his *panta rhei* argument, *everything flows*, identifying the fact that one cannot immerse in the same river twice. Immersing in it once again it is not exactly the same river – and perhaps the person descending in the water is not even the same person.

It is therefore a great advantage for the educator to be able to monitor the progress of the learner in terms of knowledge throughout the learning process [2]. Sadly, this may often be difficult as traditional testing may disturb and interfere with the consistency of the planned learning process. Another interesting question is what kind of data is in fact beneficial for the educator in the course of teaching?

The aim of this paper is to discuss these issues and suggest a way of gathering feedback from the learner throughout the temporal learning process. Such a gathering is exemplified through using The Conceptual Pond assessment application [3] as a tool to aid the teacher in making intermediary, intuitive assessments in the course of a learning experience. Two specific use cases are presented and discussed.

At the same time it is obvious that the possibility for interim assessment in realtime, may influence the way in which teaching is prepared and performed, and what teaching aids may be helpful. Interactivity obviously is only of benefit to the learning experience if the schedule and/or content may in fact be adjusted according to the new knowledge of the educator. This will be discussed on the theoretical basis of the classic five *canons* of rhetoric [4].

Some important concepts and terms must be defined:

- Learning event or similar refers to a single session of learning or a short sequence leaving no temporal space for traditional assessment.
- Realtime assessment refers to the gathering of input with no notable delay (< 3 minutes or continuous).
- Teaching refers to any kind of educational practice involving a predefined timespan of learner attention.
- Feedback refers to any kind of input or expression conveying and exposing the interests and attitudes of the learner to the teacher.
- No distinction is made between learning / education or teacher / educator.

This paper is structured as follows: after the introduction (1) important aspects of the learning situation are presented (II) using a framework of rhetorical *aptum*. Subsequently, the element of feedback is discussed (III) leading to an analysis of the leaning situation implementing *kairos* theory (IV). The continuous learning feedback application is described (V) and two use cases are presented (VI) leading to reflections on consequences for teaching (VII) and a closing section (VIII) with concluding remarks and suggestions for further study.

II. Shaping the Learning Situation

In the preparation phase of a learning event, multiple aspects are obviously taken into account by any educator. Referring to the classic *aptum* model (Figure 1.) for the appropriate formation of a communicative event may be illustrative in creating overview and suggesting terminology.

In this model, the *orator* is the transmitter of a message, in casu the teacher. Scena is the audience. *Situatio* depicts the communicative situation. *Res* is the matter or content of the communicative act. *Verba* is the actual expression or eloquent wording of the message. Though this model originates in pre-digital history it is suggested as a helpful model as it contains basic elements in communication [5] [6].

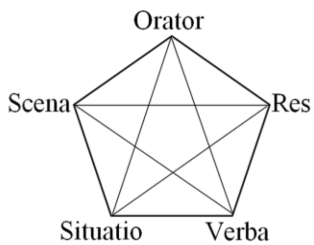


Figure 1. The aptum model

In the pentagram model, there are two relations relevant to this paper: first, it is the relation between *orator* and *scena* and, second, the relation between *res* and *verba*. In the classic rhetorical situation of the persuasive, political, or judicial speech, the five aspects are generally fixed at the beginning of the communicative act. The sequence of the speech requires this. On the other hand, a sudden change in the expression of the *scena* may at one time influence the *orator* into altering the relationship between *res* and *verba* reformulating and reshaping the speech. This interdependency between the *aptum partes* allow for a dynamic development in the communicative processes.

Though rhetoric provides a helpful framework for the understanding of communicative processes, it is not a theory originally designed directly for learning purposes. Therefore, learning taxonomies are also an important part of understanding assessment and feedback [7].

The general structure of these taxonomies pictures an increasing level of knowledge, engagement, and reflection. This progress must be verified, as suggested in the shape of *Structure of Observed Learning Outcomes* (SOLO taxonomy) [8] leading the learner on the path towards increasing complexity from the *pre-structural* level through *unistructural* and *multistructural* levels to the *relational* level, where ownership of the learning process is achieved and further leading to an *extended abstract* level.

Though these steps may not be implemented fully in the paper, the progression of complexity, ownership and engagement is worth noting as goals for learning processes.

III. Feedback

Obviously, feedback has always been part of an educational process. In primary school, a teacher will most often ask for feedback from pupils in the form of raising their hands, selected pupils making a remark or in other ways making sure that the actual teaching makes sense for the recipients. In this broad understanding, the concept of interactivity in education has always been implemented.

Though this process of feedback, interactivity, or user input is unproblematic in a manageable physical learning environment, this is not the case in other contexts. In a learning scenario with 200 students the complexity of receiving traditional feedback from so many actors would create immense cognitive stress. At the same time, lectures would be prolonged indefinitely if each student should be allowed and encouraged to interact with the lecturer.

Since the reduction of complexity is necessary, approaches from the sustainable primary school example have often been implemented: The raising of hands or the comments from a small number of students that now represent a quite limited percentage of the students. At the same time, it may be questionable whether these students actually do represent the general interests and concerns of the collective of students. Psychological factors may discourage some students from raising issues in plenum and others may want to show off their understanding in order to impress the lecturer or others.

Another unwanted aspect of collecting feedback during a learning experience is the distraction caused by the process. Cognitive stress is added to the students, as well as the lecturer having to reflect on comments that may not be relevant to the subject per se and the factor of time and sequence is always in play.

For these reasons, assessment tools such as questionnaires are often implemented to create an informed overview of the challenges of the learning situation. This feedback may be of considerable value and relevance, yet it does not always contribute in time.

IV. A Kairos Approach

Consistent with the implementation of rhetorical theory in the description of the learning situation some important issues are discussed here on the basis of rhetoric, in casu *kairos*. The concept of *kairos* designates three aspects of reaching an opportune situation for a specific task. The *kairos* triad consists of opportune time, the opportune location, and the opportune manner, as shown in figure 2:

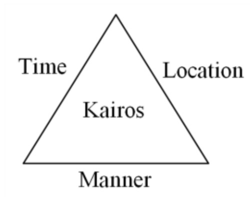


Figure 2. Kairos triad

Implementing the *kairos* model on the discussion on feedback adds methodological overview. A sustainable use of gathering and processing of feedback should be consistent with all three parameters present in this model.

In a persuasive technology discourse, B.J. Fogg argues that "the dynamics underlying suggestion technology date back to a principle of persuasion called Kairos. Discussed by ancient Greek rhetoricians, kairos means finding the opportune moment to present your message." [9]. In the use of a persuasive application for assessment, the implementation of rhetorical theory is thus not unheard of.

A. The opportune time:

- Feedback should be collected and assessment made at a moment where this does not conflict or interfere with the primary teaching task.
- Furthermore, the collected data and the new insights should be available for the teacher at the opportune moment in the learning experience.
- The collection and processing should be completed in realtime.

B. The opportune location:

- Feedback should be collected and processed in a way that does not require a physical absence from the original learning situation.
- Collecting feedback onsite should not disturb the overall concentration of the learning experience by disturbing the relevant learning scenarios.

C. The opportune manner:

Feedback should be collected and processed in such a way that:

- It is easily collected without increasing the cognitive load of the informer.
- It is easily handled and processed by the teacher.
- It allows a variety of expressions.
- It does not limit too strictly the freedom of the user to express opinions and interests not foreseen by the teacher.
- It supports anonymity as well as personal or group identification where this is relevant.
- It presents a relevant balance between detail and overview.
- It allows for storing, sharing and cross tabulation, where applicable.

Making this *kairos* analysis the foundation for the development process supports the formulation of relevant criteria for a sustainable system for the collection of feedback.

V. CONTINOUS FEEDBACK

In coherence with the analysis above and the parameters important for the collection of feedback and the facilitation of suitable assessment, it is clear that the temporal factor is of definite importance. Approaching a pedagogical implementation of interactivity may be possible only in a realtime or near-realtime context.

This may imply that the collection of feedback should not be understood as an isolated item in a sequential learning experience but rather as a underlying process, supporting the basic learning strategy. Thus, a system facilitating continuous feedback should be implemented.

In the light of experiences with technology enhanced learning and TEL-systems (Technology Enhanced Learning), it appears evident that such a continuous feedback support system should be provided through implementing a suitable assessment system capable of meetings the general demands of the *kairos* analysis.

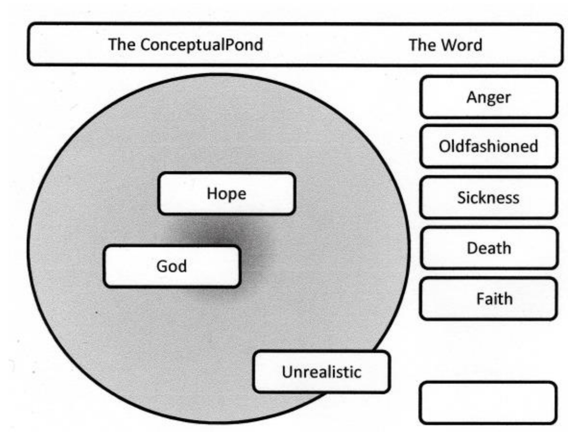


Figure 3. The Conceptual Pond – reduced in complexity

In the case studies of this paper, an intuitive interface for assessment and collection of data The Conceptual Pond application has been used. The Conceptual Pond [10] [11] is a visual interface facilitating easy and intuitive collection of qualitative data. The user interface is quite simple and allows the user to select one or several words or phrases that resemble the intended expression of the user. If no suggested text is fitting, the users may add their own text. A purely visual representation through images is also supported. After selecting the relevant text, the user applies this text to a circle marking the distance from the center which corresponds to which the user finds this textual expression important.

By doing so, the user easily supplies feedback, even on quite complex questions, as exemplified in Case Study 1. Once the qualitative data are collected through using the system, the collected data are instantly quantified into graphs creating an instant overview for the teacher. Using this system supports the desire for continuous feedback, as the user can always move or substitute the text with something else. Such changes are immediately reflected in the graphs available for the teacher in realtime.

Though the overall argument of this paper is not tied to the implementation of any one system for interactive assessment (such as The Conceptual Pond), the facilities of the system is important as the system should satisfy all or most of the criteria listed in relation to the *kairos* analysis. Obviously, ease of use, user freedom, and instant teacher overview are important factors.

Referring to the principles of the hermeneutic circle of Gadamer [12], it is obvious that the continuous feedback approach, in reality resembles the process of recognition as devised in this model. Issues important to the user in the beginning of the learning experience may lose importance or be transformed into new questions or issues of relevance. Through a continuous feedback approach both learner and teacher should benefit from the pedagogical potential of this.

VI. USE CASES

Two tests using The Conceptual Pond for advanced feedback are presented in this paper. In both cases, the focus has been on teenagers, but they are not part of the same group. They have not been chosen for the tests through any special selection process.

A. Reflections on the movie The Word

A group of 23 Danish teenagers from confirmation class (13-14 years old) viewed a selected scene from the play "The Word" by Danish playwright Kaj Munk (1962) in a well-known 1955 movie adaptation of Carl Th. Dreyer. The test took place in 2012.

Despite an antiquated visual language with remarkably slow dialogue and black/white aesthetics, the scene of the resurrection of a woman deceased in childbirth appeared quite moving for the teenagers.

Immediately after the screening of the film and before proceeding with the scheduled program, all teenagers entered their immediate impression and thoughts to the assessment interface. In this way, feedback was secured in a very short time and answers were stored as well as quantified and available instantly for the teacher as a graph.

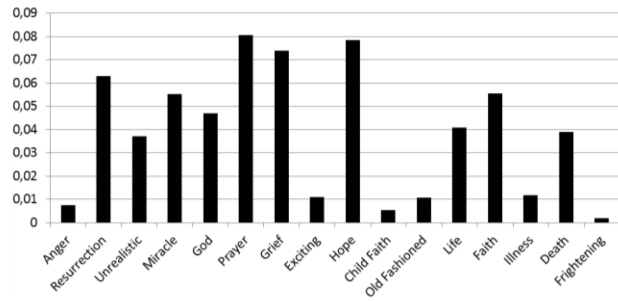


Figure 4. Graphic presentation

Completing the quite complex feedback process on computers was performed in an average of 105 seconds (variation 75 – 135 seconds). In a subsequent testing under similar conditions and a similar use case (2014) feedback was even more rapid using a touch device (iPad) reducing average feedback time to 95 seconds (variation 60 – 110 seconds). This observation suggests that the more tangible handling offers enhanced intuitiveness in interaction with the application and therefore greater speed.

The graph (figure 4) was available for the teacher immediately after the gathering of information, thus empowering him immediately to focus the teaching program in the direction of issues important to the learners.

From a pedagogical perspective, the feedback was rich for several reasons:

- All teenagers made a relevant contribution and they were heard equally. In ordinary teaching conditions in class, it would only be possible to hear the impressions of a selected number of students. Now, everyone did participate in a synchronous sharing of thoughts.
- Some of the results were surprising and provided a good foundation for designing the further teaching process. For example, the teenagers did not mind much that the movie was "old-fashioned" and the vivid scenes of a woman rising from the dead did not frighten them. "Prayer" and "hope" have scores twice as high as the quite relevant "unrealistic", suggesting that the students did internalize the intended message of the author rather than reacting to the surface level of the play.
- The teenagers were able to compare their own reflections with the reflections of peers, which is an important persuasive factor, as suggested by Oinas-Kukkonen and Harjumaa [13].

In a subsequent inquiry, all teenagers preferred The Conceptual Pond to traditional aids for assessment such as questionnaires and multiple choice tests. Such assessment tools are regularly used in school, so the teenagers have a considerable familiarity using them.

At the same time, a group discussion was implemented in order to verify that the expressions codified in the data-set were representative of the expressions uttered in a group discussion. This was the case.

B. Feedback at teaching on robots and ethics

A group of 24 Danish teenagers from the municipal school of Nørager (13-14 years old) were taught on the subject of robots and ethics. The test took place in 2013.

The Conceptual Pond was used twice during the actual teaching process, making the teenagers multitask in order to listen, reflect, and answer synchronously. For this reason, the time used for filling in the pond is not specified. It was done synchronously with teaching.

The first question asked the students which one of the presented subjects was the most interesting to them. Not surprisingly, “dangerous robots” as well as “cyborgs” were the most interesting and “ethics” scored quite low. “Talking with a robot” was also quite low. In the course of teaching, the students were encouraged to interact with a chat-robot providing them with an initial experience of communicating meaningfully with an Artificial Intelligence-system [14]. After this unexpected experience, the subject “talking to a robot” advanced to being the second most interesting subject, almost tripling in popularity.

From a pedagogical perspective the feedback was rich for several reasons:

- The feedback allowed the teacher to focus on subjects relevant to the learner. Obviously, “dangerous robots” could not be the full content of the learning experience. But the sequence in which subjects were presented was influenced and “dangerous robots” often used as examples.
- Through the continuous feedback, the teacher was able to detect the growing fascination by “talk to a robot” end elaborate on this in the learning experience. If the initial expressions of interest had not been adjusted through continuous feedback, the teacher might have focused too narrowly on “dangerous robots” reducing the potential of teaching to fit the changing interests of the students.

Later in the same teaching session, the teenagers were challenged to sort different robots in the order in which they found them the scariest. The objective was to confirm or falsify the assumption that the Uncanny Valley observation of Masahiro Mori [15] also applies to Danish teenagers.

After filling out The Conceptual Pond the teenagers watched a video presenting the Uncanny Valley. Much to their surprise, the graph presenting their answers in detail did correspond exactly with the uncanny Valley argument.

From a pedagogical perspective, the feedback was rich as it revealed not only knowledge, but also interest and immersion and facilitated an enhanced level of reflection not typical of students at this level of education:

- The teenagers instantly understood themselves as part of a scientific experiment. In this way they came to value scientific tradition, suddenly realizing research was relevant.
- In contrast to a traditional test on paper that would require manual processing, the result was immediately available and presented visually on a big screen. The three aspects of *kairos* are thus implemented.
- The emotional impact of peer comparison in relation to the reception of robots may be expected to support the reflection on the themes presented in the teaching. The students were genuinely surprised realizing that they did indeed mirror the Uncanny Valley theory of Mori.

From these two quite different use cases, it may be derived with a substantial degree of credibility that continuous feedback is a strong persuasive factor and it does add extra possibilities, and, therefore, extra quality to traditional teaching.

No difficulties were detected through the tests except for the fact, that links to the online environment of The Conceptual Pond should be distributed in advance. Since students in Danish schools use a variety of devices from smartphones to iPad’s, to laptops, it is essential that the feedback system supports all platforms. This is one of the advantages of The Conceptual Pond in comparison to a number of TEL applications designed with a limitation to a specific platform.

VII. CONSEQUENCES FOR TEACHING

Implementing feedback in a rather radical way, as it is suggested and exemplified above, does pose some challenges to traditional teaching. The content of teaching may mostly remain the same, but the sequence and context in which it is shared may be required to change. The *kairos* approach influences education in double movement:

Continuous learning feedback, on one side, allows the educator to be informed about the position, knowledge, or attitude of the learner. On the other hand, it simultaneously offers the teacher the opportunity of sharing this information with the learner in a real-time environment.

Referring again to classical rhetoric the five *canons* for preparing a piece of communication, this model may act as a framework or example. The five canons consist of *inventio* (developing and refining arguments and other content), *dispositio* (arranging and organizing content), *elocutio* (stylistic choices and eloquence), *memoria* (memorizing), and *actio* (delivery). From a traditional point of view, this process is sequential. The preparation takes place in steps in the right order. Implementing continuous feedback and taking advantage of the possibilities may require a more liquid approach to the rhetorical situation (in casu teaching). The sequence prepared in *dispositio* may be challenged which may require the educator to focus more directly on *inventio* and *eloqutio* recognizing that the teaching prepared may in fact have to be changed at some level adjusting to the enhanced level of knowledge of the learners facilitated by structured feedback.

VIII. Conclusion and Future Research

In terms of theory, there is a substantial leap from classical rhetoric to the implementation of a continuous feedback system through a digital application, as is analysed and suggested in this paper. Nevertheless, this leap is important as the implementation of continuous learning feedback appears to be promising in developing novel concepts for learning.

Though traditional methods for assessment and acquiring feedback such as written tests, oral consultation, questionnaires, and multiple choice tests may be helpful in a number of contexts, it is clear that at least two challenges remain. One challenge is the element of *kairos*: Acquiring the right knowledge at the right time – and even sharing it with the learner simultaneously. Another challenge is the observation that feedback may be more comprehensive than traditional assessment. An example could be the suggestions made [16], that in the field of e.g. cultural mediation the predominantly simplified nature of traditional assessment is inadequate for monitoring the complex experiences and reflections done by the learner.

Much research is left to be done in the field of using feedback in a digital system. The aspect of ownership to the educational process if the user is engaged more in reflecting and commenting openly should be explored deeply. The use of a familiar electronic device such as your own smartphone instead of potentially alienating white paper may also be of importance.

In this paper, some of the possible uses of such system have been described. There are many other use cases to explore and recognitions to be made from these, again adding to agile development process of continuous learning feedback.

References

[1] R. Riding and S. Rayner, ”Cognitive Styles and Learning Strategies.”, London 2000.

[2] B. Bloom, “Taxonomy of Educational Objectives, the classification of educational goals – Handbook I: Cognitive Domain,” New York, 1956.

[3] C. Grund Sørensen and M. Grund Sørensen, “The Conceptual Pond – A Persuasive Tool for Quantifiable Qualitative Assessment” in K. Blashki and P. Isias, “Emerging Research and Trends in Interactivity and the Human - Computer Interface”, pp.449-469, IGI Global, 2013.

[4] E.P.J. Corbett and R.J. Connors, “Classical Rhetoric for the Modern Student”, New York 1999.

[5] P. Hasle.P. and A.K.K. Christensen, “Classical Rhetoric and a Limit to Persuasion,” Persuasive 2007, LNCS 4744, Springer, 2007.

[6] P. Hasle, “The Persuasive Expansion - Rhetoric, Information Architecture, and Conceptual Structure,” in ICCS 2006, Editors H. Schärfe, P. Hitzler, and P. Øhrstrøm, Springer, 2006.

[7] B. Bloom, “Taxonomy of Educational Objectives, the classification of educational goals – Handbook I: Cognitive Domain,” New York, 1956.

[8] J. Biggs and C. Tang, “Teaching for Quality Learning at University,” (3.rd ed.), Buckingham, SRHE and Open University Press, 2007.

[9] B.J. Fogg, “Persuasive Technology – Using Computers to Change what we Say and Do”, San Francisco, 2003, p.41

[10] C. Grund Sørensen and M. Grund Sørensen, “The Conceptual Pond – A Persuasive Tool for Quantifiable Qualitative Assessment” in K. Blashki and P. Isias, “Emerging Research and Trends in Interactivity and the Human - Computer Interface”, pp.449-469, IGI Global, 2013.

[11] C. Grund Sørensen, “Intuitive Surveying and Quantification of Qualitative Input through the Conceptual Pond”, in R. Behringer and G. Sinclair, “IWEPLET 2013”, pp. 135-142, 2013.

[12] H.G. Gadamer, “Truth and Method”, Bloomsbury Academic 2004

[13] H. Oinas-Kukkonen, H.and M. Harjumaa, “Persuasive Systems Design: Key Issues, Process Model, and System Features”, in “ Communications of the Association for Information Systems 2009/24”, pp. 485-500 2009.

[14] Http:/www.alice.pandorabots.com/

[15] Y. Mori, “The Uncanny Valley”, Energy 7, pp.33-35, 1970.

[16] E. Hooper-Greenhil,“Measuring Learning Outcomes in Museums, Archives and Libraries: The Learning Impact Research Project”, International Journal of Heritage Studies Vol. 10, Issue 2, 2004.

Interpreting Psychophysiological States Using Unobtrusive Wearable Sensors in Virtual Reality

Alberto Betella, Daniel Pacheco, Riccardo Zucca, Xerxes D. Arsiwalla, Pedro Omedas
SPECS, NRAS, Universitat Pompeu Fabra
Barcelona, Spain
Email: alberto.betella@upf.edu

Antonio Lanatà, Daniele Mazzei, Alessandro Tognetti, Alberto Greco, Nicola Carbonaro
Research Centre "E. Piaggio", University of Pisa
Pisa, Italy

Johannes Wagner, Florian Lingenfelser
Human Centered Multimedia, University of Augsburg
Augsburg, Germany

Elisabeth André
Human Centered Multimedia, University of Augsburg
Augsburg, Germany

Danilo De Rossi
Research Centre "E. Piaggio", University of Pisa
Information Engineering Department, University of Pisa
Pisa, Italy

Paul F.M.J. Verschure
SPECS, NRAS, Universitat Pompeu Fabra
ICREA, Institució Catalana de Recerca i Estudis Avançats
Barcelona, Spain

***Abstract*—One of the main challenges in the study of human behavior is to quantitatively assess the participants' affective states by measuring their psychophysiological signals in ecologically valid conditions. The quality of the acquired data, in fact, is often poor due to artifacts generated by natural interactions such as full body movements and gestures. We created a technology to address this problem. We enhanced the eXperience Induction Machine (XIM), an immersive space we built to conduct experiments on human behavior, with unobtrusive wearable sensors that measure electrocardiogram, breathing rate and electrodermal response. We conducted an empirical validation where participants wearing these sensors were free to move in the XIM space while exposed to a series of visual stimuli taken from the International Affective Picture System (IAPS). Our main result consists in the quantitative estimation of the arousal range of the affective stimuli through the analysis of participants' psychophysiological states. Taken together, our findings show that the XIM constitutes a novel tool to study human behavior in life-like conditions.**

Keywords-Affect analysis, affective states, ecological validity, EDR, HRV, wearable sensors, XIM

I. Introduction

In the last two decades, the advances in virtual reality techniques led to their application to a wide range of scientific fields. In particular, virtual reality has been generating important contributions to scientific research related to the understanding of human behavior [1]. Virtual reality systems, in fact, allow to setup ecologically valid environments where user can act and behave in life-like conditions. At the same time, researchers can have a systematic control of the stimuli presented.

The empirical validation of users' subjective experience in such environments is commonly conducted through the administration of self-assessment questionnaires. Besides this, the users' behavioral patterns (e.g., spatial location, gaze, etc.) can be recorded and analyzed, thus offering useful insights on users' explicit and implicit behavior.

In recent years, due to the improvements in hardware portability, an important addition to the study of human behavior is the measure of psychophysiological signals such as electrodermal response (EDR) and electrocardiogram (ECG). Past literature has shown that these measures are directly related to the Autonomic Nervous System (ANS) and provide means to quantitatively assess a number of human affective states [2], [3], [4].

However, one of the common issues when recording these measures in ecologically valid conditions, is the quality of the signal. As a matter of fact, full body movement and

 ISBN: 978-1-61208-325-4

embodied interaction (e.g., natural gestures) often result in a series of artifacts in the signal that don't allow the extraction of features that assess the users' psychophysiological state. For this reason, there is still a tendency to measure these signals only in conventional lab conditions where the user is sitting or has a limited range of movements and interaction.

To address this problem, we enhanced the eXperience Induction Machine (XIM), an immersive space we constructed to conduct experiments on human behavior, with unobtrusive wearable sensors capable of measuring users' psychophysiological signals [5]. Through this latest addition, we obtained a general-purpose infrastructure to support a broad range of behavioral studies that include human affective states.

We validated our new infrastructure by conducting an experiment where participants were free to move or perform gestures in XIM and, at the same time, were exposed to a series of visual stimuli taken from the International Affective Picture System (IAPS) [6], while recording their ECG and EDR in real-time.

Our results show that we are able to distinguish the arousal range of the stimuli presented to the participants by measuring their psychophysiological signals.

II. Materials and Methods

A. The eXperience Induction Machine

The XIM (Figure 2) is an immersive space constructed to conduct empirical studies on human behavior in ecologically valid situations that involve full body interaction [7].

The XIM covers an area of about 25 square meters and is equipped with a number of effectors that include 8 projectors, 4 projection screens, a luminous interactive floor and a sonification system.

Along with the effectors, the XIM features a pool of sensors to measure users' explicit and implicit behavior, including a marker-free multi-modal tracking system, floor-based pressure sensors and microphones. Next to these, we added wireless and wearable devices capable of measuring body posture, arm orientation, hand position and fingers movements in real-time, as well as the following psychophysiological signals: ECG, EDR and breathing rate (BR).

In the design of the system, these implicit signals were specifically selected because of their reliability in terms of inference of affective states and because the hardware requirements allowed to mount the sensors on tiny wearable devices.

To provide an ecological form of interaction, the wearable devices were integrated into two main interfaces: a) a sensing glove for the simultaneous acquisition of hand gestures and EDR and b) a sensing shirt for the acquisition of ECG and respiration.

1) The sensing shirt and glove: On the one hand, the sensing shirt (Smartex srl, Italy) is used in the XIM to acquire ECG, BR and triaxial accelerometric data [8]. This device has been adopted in previous studies on long-term monitoring of chronic patients, focusing on the early prevention of cardiovascular diseases [9]. The sensing shirt features a front pocket that contains a tiny electronic battery-powered unit that streams the acquired data through a bluetooth connection.

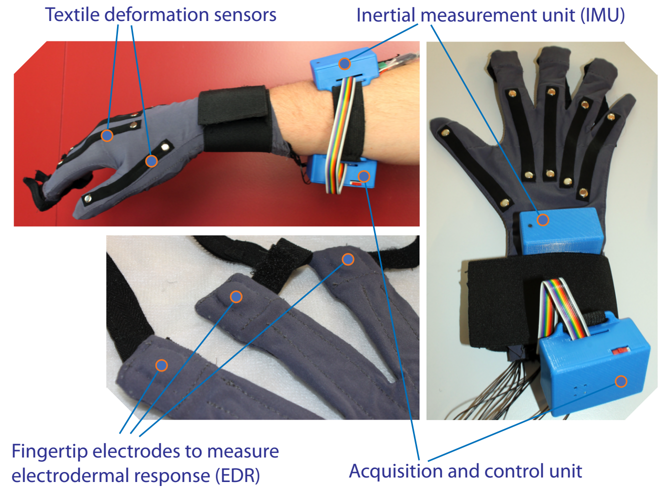


Figure 1. The sensing glove and its main components. See text for further explanation.

On the other hand, the multi-parameter sensing glove (Figure 1) was specifically designed for the XIM space. It measures both explicit and implicit signals (forearm orientation, fingers position and EDR) on a single hand.

In previous work, we demonstrated that textile electrodes are comparable in terms of bio-electric properties to standard Ag/AgCl electrodes [10]. For this reason, in our sensing glove we integrated textile electrodes placed at the fingertips to measure EDR.

In addition, the use of a textile system presents a series of advantages in terms of portability and usability for long-term monitoring, and provides minimal constraints in terms of natural movements and gestures.

Finger motion tracking is obtained through five textile deformation sensors that have been specifically designed and located on the metacarpo-phalangeal hand joints (Figure 1). These sensors, previously used for respiration monitoring in [11], are made of knitted piezoresistive fabric (KPF) material (Smartex srl, Italy). Finger movements produce local deformations in the fabric thus modifying the electrical resistance of the sensors. This resistance is highly correlated with the single finger degree of flexion. The sensor signal is characterized by a slow baseline drift due to the intrinsic characteristics of textile substrate, for this reason a custom-built algorithm for hand gesture recognition was developed [12].

Both EDR and deformation signals are acquired and elaborated in real-time by using a dedicated wearable and wireless electronic unit. Moreover, forearm orientation is measured by an Inertial Measurement Unit (HMC6343, Honeywell, MN, USA) embedded in the glove's electronics and placed on the dorsal area of the forearm, close to the wrist.

2) The sensing platform: The data coming from the wearable devices is sent to the XIM sensing platform which is responsible for capturing and processing in real-time raw sensor data [5]. This platform is implemented using the Social Signal Interpretation (SSI) framework [13].

The data stream of each sensor is transmitted through a dedicated channel and preprocessed (e.g., forearm orientation, fingers position and EDR from the sensing glove are assigned to separate channels).

The sensing platform synchronizes the incoming streams by establishing a stable connection with all the sensors and by eventually starting to buffer the data streams. Each buffer is compared upon regular time intervals according to an internal timestamp and synchronized if necessary. After the synchronization, each one of the signals is processed individually to separate meaningful information from artifacts.

B. Empirical Validation

1) Selection of the stimuli: 12 visual stimuli were chosen from the pictures of the IAPS collection [6], each of which was representing a different rating value of arousal, thus covering the entire scale of arousal from a minimum rating of 1.72 to a maximum of 7.34, as shown in Table I.

2) Sample and Protocol: A total of 7 Subjects (4 females, mean age 29.7, SD±3.9) participated in the empirical validation.

Prior to the exposure, the participants were helped to wear the sensing shirt and the glove to measure ECG and EDR respectively. A short phase of connection testing followed.

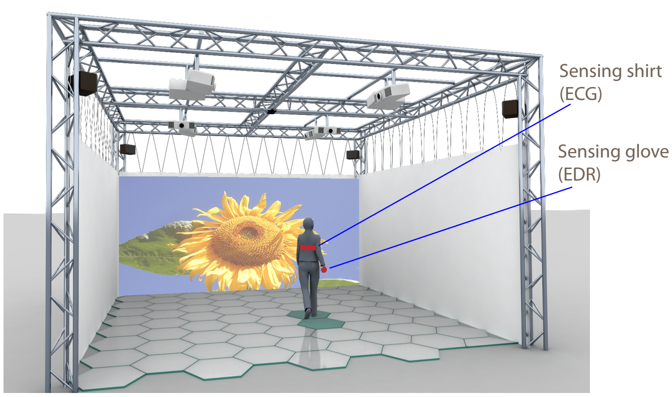


Figure 2. Schematic illustration showing the experimental setting in the eXperience Induction Machine (XIM).

The participants were then instructed to enter the XIM and place themselves at the designated starting point in the center of the room.

A 5-minutes baseline recording phase followed where the participants were asked to maintain a natural standing position and relax as much as possible.

After the baseline recording, the first affective stimulus was displayed on the main screen of the XIM. The order of presentation of the stimuli was randomized for each experimental session.

Each one of the 12 stimuli was displayed for at least 20 seconds, after which a "beep" sound followed to notify the user about the possibility to switch to the following stimulus. To do so, the user was instructed to close all the fingers of the hand wearing the glove, hence generating a "grabbing" event. This event was sent to the sensing platform to provide an accurate

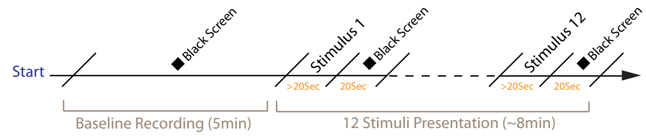


Figure 3. Schematic timeline of a single experimental session.

time marker for each stimulus. A 20-seconds black screen was presented in-between each stimulus (Figure 3).

During the entire experimental session, participants were asked to maintain a standing posture and were free to move when exposed to the visual stimuli (pictures from the IAPS collection) displayed on the main XIM screen (Figure 2).

C. Acquired Signals

1) Electrocardiogram: The ECG signal was used to extract the Heart Rate Variability (HRV) [14], which indicates the time variation between two consecutive R-peaks. Previous research has shown that HRV is directly related to sympatho-vagal balance [2]. To extract HRV, the ECG was pre-filtered using a Moving Average Filter (MAF) to extract and subtract the baseline.

The Heart Rate (HR) is $HR = \frac{60}{t_{R-R}}$, where t_{R-R} is the time interval between two successive R-peaks: consequently, we treated the ECG signal by detecting the QRS complexes through the automatic QRS detection algorithm proposed by Pan-Tompkins [15] and by extracting R-peaks. The obtained HR resulted in a time series sequence of non-uniform RR intervals, hence it was re-sampled in accordance to the recommendations of Berger et al. [16].

2) Electrodermal Response: EDR is obtained as the ratio between an imposed continuous voltage of 0.5 V applied to two fingers and flowing current.

EDR is related to variations in the electrical conductance of the skin, due to the changes in the sweat gland activity [17] after a sympathetic stimulation. For this reason hereinafter we refer to EDR as Skin Conductance (SC).

Since sweat gland activity is controlled by the Sympathetic Nerve Activity (SNA) [3], [4], SC constitutes a good measure to monitor changes in the Autonomic Nervous System (ANS). This has been validated in previous studies showing that electrodermal changes are associated to the arousal of the stimuli [18], [19].

The SC signal is characterized by a tonic and a phasic component: the Skin Conductance Level (SCL) and the Skin Conductance Response (SCR) respectively.

On the one hand, SCL is the slowly varying baseline level of skin conductance, related to an overall state of the body. On the other hand, SCR arises within a predefined response window of $1-5$ seconds after a given stimulus and is directly related to it [20].

Consecutive SCRs can result in an overlap when the inter-stimulus time interval is shorter than SCR recovery time. To address this issue, in this study we adopted a deconvolution model that allows the direct estimation of the SudoMotor

 ISBN: 978-1-61208-325-4

TABLE I. SELECTION OF THE 12 STIMULI FROM THE INTERNATIONAL AFFECTIVE PICTURE SYSTEM (IAPS) DATABASE [6].

ID	IAPS Catalog ID	Description	Arousal mean(SD)	Subset α	Subset β	Subset γ
1	7175	Lamp	1.72($\pm$1.26)	A_1	A_1	A_1
2	7020	Fan	2.17($\pm$1.71)	A_1	A_1	A_1
3	5030	Flower	2.74($\pm$2.13)	A_1	A_1	A_1
4	7547	Bridge	3.18($\pm$2.01)	A_1	-	A_1
5	7512	Chess	3.72($\pm$2.07)	A_1	-	A_2
6	9280	Smoke	4.26($\pm$2.44)	-	-	A_2
7	9171	Fisherman	4.72($\pm$2.17)	-	-	A_2
8	9582	Dental Exam	5.29($\pm$2.21)	A_2	-	A_2
9	9611	Plane Crash	5.75($\pm$2.44)	A_2	-	A_3
10	9622	Jet	6.26($\pm$1.98)	A_2	A_2	A_3
11	9412	Dead Man	6.72($\pm$2.07)	A_2	A_2	A_3
12	3000	Mutilation	7.34($\pm$2.27)	A_2	A_2	A_3

Nerve Activity (SMNA) which is the controller of the eccrine sweat glands activity [21]. The SMNA shows an inter-stimulus time interval shorter than the SCR signal thus avoids the overlapping problem. Specifically, the SCR is the result of a convolution model between SMNA and the following biexponential Impulse Response Function (IRF), which is called Bateman function [22]:

$$IRF(t) = (e^{-\frac{t}{\tau_1}} - e^{-\frac{t}{\tau_2}}) \cdot u(t) \tag{1}$$

where u(t) is the step function. This function is characterized by a rapid increase and a slower recovery. More specifically, $\tau 1$ is a time constant that describes the rapid slope, taking information on time course of the evacuation of sweat from the compartment model, while $\tau 2$ is the related slow recovery slope.

D. Physiological features

1) Heart Rate Variability: HRV is related to the time intervals between heartbeats [14].

A number of features were extracted in both the time and the frequency domains. In the time domain, we extracted statistical parameters and morphological indices. We fixed a time window (NN) and extracted the following features: a) simple MNN and SDNN, which correspond to the mean value and to the standard deviation of the NN intervals, respectively, b) the root mean square of successive differences of intervals (RMSSD) and c) the number of successive differences of intervals which differ by more than 50 ms (pNN50 % expressed as a percentage of the total number of heartbeats analyzed).

In the HRV frequency domain analysis, three main spectral components were distinguished in a spectrum calculated from short-term recordings: Very Low Frequency (VLF), Low Frequency (LF), and High Frequency (HF) components. Short terms recordings are intended as the time duration of HRV signal segments. In this work HRV segments were in agreement with picture presentation time. Current HRV research in the frequency domain analysis suggests that even though the frequency band division represents a unique, non-invasive tool to achieve an assessment of autonomic function, the use of HF and LF components does not allow to precisely assess the state of sympathetic activation. Therefore, along with the estimation of the Power Spectral Density in the VLF, LF, and HF band, we also calculated the LF/HF PSD Ratio which provides information about the Sympatho-Vagal balance [23] (Table II).

2) Skin Conductance: SC decomposition to its components was performed using Ledalab software package in MATLAB [24]. The signal was filtered by means of a low pass zero-phase forward and reverse digital filter [25] with a cutoff frequency of 2 Hz.

The phasic features were calculated within a time window (response window) of 5 seconds length after stimulus onset. We extracted the number of SCRs within the response window (nSCR), the latency of the first SCR (Lat), the Amplitude-Sum of SCRs (reconvolved from phasic driver-peaks) (AmpSum), the average phasic driver activity (Mean.SCR) (time integral over response window by size of response window), the variance of the phasic driver signal (Var.SCR), the Phasic driver area under curve (AUC.SCR) and the max phasic driver amplitude (Max.SCR).

From the tonic driver signal we extracted the following features: average level of (decomposed) tonic component (Mean.Tonic), variance of the tonic driver signal (Var.Tonic) and number of the non-specific response (i.e., the spontaneous skin conductance response unrelated to a specific stimulus) (NSR) (Table II).

TABLE II. FEATURES EXTRACTED FROM HRV AND SC.

Feature set	Signals
MNN, SDNN, RMSSD, pNN50, VLF, LF, HF, LF/HF	HRV
Lat, nSCR, Mean.SCR, Var.SCR, Max.SCR AmpSum, AUC.SCR, Mean.Tonic, Var.Tonic, NSR	SC

E. Data Analysis

A number of fixed windows were used to segment the signals (EDR, HRV) in accordance to the experimental protocol (Section II-B2). On the one hand, to compute each feature of the phasic component, the EDR signal was segmented using a 5 seconds window starting after the presentation of each black screen that anticipates the IAPS image. On the other hand, longer windows of 20 seconds length (i.e., the entire duration of each one of the 12 stimuli) were used to compute the HRV parameters.

The extracted features (Table II) were normalized in accordance to the baseline recording. Subsequently, the dataset was divided into 3 different subsets as follows:

- α: A_1 refers to arousal intervals 1-3, A_2 refers to arousal intervals 5-7. Each class comprises 5 stimuli.

- β: A_1 refers to arousal intervals 1-2, A_2 refers to arousal intervals 6-7. Each class comprises 3 stimuli.
- γ: A_1 refers to arousal intervals 1-3, A_2 refers to arousal intervals 3-5, A_3 refers to arousal intervals 5-7. Each class comprises 4 stimuli.

As a result, the subsets α and β represent a 2-class problem, while subset γ represents a 3-class problem (Table I).

A statistical inference analysis was conducted by means of non-parametric tests due to the non-gaussianity of the sample set to verify the null-hypothesis of having no statistical difference between the classes for both the 2-class (datasets α and β) and the 3-class (dataset γ) problems.

The α and the β datasets were submitted to a Mann-Whitney test. Instead, the γ dataset was submitted to a Kruskal-Wallis (KW) test, and a subsequent pairwise comparison among A_1, A_2 and A_3, was performed by means of Mann-Whitney tests with a Bonferroni correction.

F. Pattern Recognition

The pattern recognition process aimed to investigate whether the arousal sessions could be distinguished in both the 2-class and the 3-class recognition problems.

Taking into account the entire dataset of features, the dimension of the features space was reduced through the application of the Principal Component Analysis (PCA) where we consider the number of PCs that explain 90% of the total variance.

In this work, the classification stage was performed in order to classify each sample of the dataset according to the set of classes reported in Section II-E.

We tested the following classifiers: Linear Discriminant Classifier (LDC), Quadratic Discriminant Classifier (QDC), Mixture Of Gaussian (MOG), k-Nearest Neighbor (k-NN), Kohonen Self Organizing Map (KSOM), Multilayer Perceptron (MLP) and Nearest Mean Classifier (NMC) [26], [27], [28].

Among the aforementioned classifiers, the LCD [29] showed the highest recognition accuracy and consistency in arousal discrimination. For this reason, and for the sake of brevity, here we limit the report and discussion of our results to the LCD. The performance of the classification process was examined using the confusion matrix, which represents the capacity of the algorithm to recognize each sample as belonging to one of the predefined classes. In details, a more diagonal confusion matrix corresponds to a higher degree of classification.

Here, we used a training set of 80% of the whole features dataset while the remaining 20% was used as testing set. The validity of the classification model was evaluated using computational techniques and, in particular, the cross-validation method. More specifically, we performed 40-fold cross-validation steps in order to obtain unbiased results. The final results were expressed as the mean and the standard deviation of the 40 computed confusion matrices.

TABLE III. CONFUSION MATRIX OF THE LDC CLASSIFIER FOR THE 2-CLASS PROBLEM FOR α AND β DATASETS. THE RESULTS WERE OBTAINED AFTER 40 CROSS-FOLD VALIDATIONS.

	Dataset α		Dataset β	
LDC	A_1	A_2	A_1	A_2
A_1	**87.27±6.19**	7.72±15.12	**95.36±6.77**	0.00±0.00
A_2	12.72±6.19	**92.27±15.12**	4.64±6.77	**100.0±0.00**

TABLE IV. CONFUSION MATRIX OF THE LDC CLASSIFIER FOR THE 3-CLASS PROBLEM FOR γ DATASET. THE RESULTS WERE OBTAINED AFTER 40 CROSS-FOLD VALIDATIONS.

LDC	A_1	A_2	A_3
A_1	**88.89±10.19**	2.78±6.11	5.56±10.44
A_2	7.72.56±7.45	**85.56±12.01**	21.11±13.74
A_3	3.89±5.43	11.67±11.09	**73.33±11.29**

III. RESULTS

An intra-subject analysis was performed for all the subjects and all the extracted features.

No statistically significant features were found in the α dataset, while for the β dataset we found a significant difference in the LF feature ($p < .05$) between the two classes A_1 and A_2. The Kruskal-Wallis (KW) test (Section II-E) indicated a statistical difference among the three classes A_1, A_2 and A_3 of the γ dataset. The pairwise comparison showed that both the features RMSSD and HF were significantly different in the three classes ($p < .01$).

As a result of the pattern recognition phase (Section II-F), the LDC classifier accounted for a high accuracy in the recognition of both the 2-class and the 3-class problems (Tables III and IV respectively).

IV. CONCLUSION

To address the question of how known emotional responses that have been established under standard laboratory conditions generalize to ecologically valid environments, we enhanced the eXperience Induction Machine (XIM) with unobtrusive wearable sensors: the sensing shirt and the sensing glove. These devices allow to measure human behavioral patterns (e.g., body posture, finger movements, etc.) as well as a number of psychophysiological signals (ECG, BR, EDR).

To validate our augmented infrastructure, we conducted an experiment in the XIM where participants were exposed to a series of visual stimuli taken from the IAPS collection, while wearing the sensing shirt and glove without any physical constraint (i.e., they were able to move freely in the space and to perform gestures).

Our results show that a number of HRV-extracted features (LF, RMSSD, HF) were coherent with the arousal ratings of the stimuli. Although the statistical significance obtained was limited to a small set of features, accuracy was high. A multivariate analysis with a LDC considering the entire dataset accounted for an accuracy between 73.3% and 88.9% in the 3-class problem, and exceeding 87% in the 2-class problem. An implication of our findings is the possibility to recognize users' level of arousal in the XIM with a high accuracy.

The addition of tailored-made solutions (e.g., the sensing glove) to the XIM infrastructure constitutes a novel approach when compared to similar devices to measure psychophysiological states, such as the Q Curve by Affectiva Inc. [30]. Our system as a whole offers a higher number of parameters that can be measured, as well as a higher level of integration and future extensibility.

Taken together, our results show that XIM constitutes an ideal tool to study human behavior in ecologically valid conditions by acquiring both explicit and implicit measures.

Future improvements will consist in empirical validations of our system in more complex life-like conditions (e.g., fast walking, elaborate gestures and body postures, etc.).

V. ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7-ICT-2009-5) under grant agreement n. 258749 [CEEDS]. The Generalitat de Catalunya (CUR, DIUE) and the European Social Fund are supporting this research.

REFERENCES

[1] C. J. Bohil, B. Alicea, and F. a. Biocca, "Virtual reality in neuroscience research and therapy," Nature Reviews Neuroscience, vol. 12, no. December, Nov. 2011.

[2] P. Stein, M. Bosner, R. Kleiger, and B. Conger, "Heart rate variability: a measure of cardiac autonomic tone," American Heart Journal, vol. 127, no. 5, 1994, pp. 1376–1381.

[3] D. Fowles, M. Christie, R. Edelberg, W. Grings, D. Lykken, and P. Venables, "Publication recommendations for electrodermal measurements," Psychophysiology, vol. 18, no. 3, 1981, pp. 232–239.

[4] W. Boucsein, Electrodermal activity. Springer Verlag, 2011.

[5] J. Wagner, F. Lingenfelser, E. André, D. Mazzei, A. Tognetti, A. Lanatà, D. De Rossi, A. Betella, R. Zucca, P. Omedas, and P. F. M. J. Verschure, "A sensing architecture for empathetic data systems," in Proceedings of the 4th Augmented Human International Conference, ser. AH '13. New York, NY, USA: ACM, 2013, pp. 96–99.

[6] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," The Center for Research in Psychophysiology, University of Florida, Gainesville, FL, Tech. Rep. A-8, 2008.

[7] U. Bernardet, A. Väljamäe, M. Inderbitzin, S. Wierenga, A. Mura, and P. F. M. J. Verschure, "Quantifying human subjective experience and social interaction using the eXperience Induction Machine." Brain research bulletin, vol. 85, Nov. 2010, pp. 305– 312.

[8] R. Paradiso, G. Loriga, and N. Taccini, "A wearable health care system based on knitted integrated sensors," Information Technology in Biomedicine, IEEE Transactions on, vol. 9, no. 3, 2005, pp. 337–344.

[9] E. Scilingo, A. Gemignani, R. Paradiso, N. Taccini, B. Ghelarducci, and D. De Rossi, "Performance evaluation of sensing fabrics for monitoring physiological and biomechanical variables," Information Technology in Biomedicine, IEEE Transactions on, vol. 9, no. 3, 2005, pp. 345–352.

[10] A. Lanatà, G. Valenza, and E. Scilingo, "A novel EDA glove based on textile-integrated electrodes for affective computing," Medical & Biological Engineering & Computing, vol. 50, no. 11, 2012, pp. 1163–1172.

[11] R. Paradiso and L. Caldani, "Electronic textile platforms for monitoring in a natural environment," Research Journal of Textile and Apparel, vol. 14, no. 4, 2010.

[12] N. Carbonaro, A. Greco, G. Anania, G. Dalle Mura, A. Tognetti, E. P. Scilingo, D. De Rossi, and A. Lanatà, "Unobtrusive Physiological and Gesture Wearable Acquisition System: A Preliminary Study on Behavioral and Emotional Correlations," in GLOBAL HEALTH 2012, The First International Conference on Global Health Challenges, 2012, pp. 88–92.

[13] J. Wagner, F. Lingenfelser, T. Baur, I. Damian, F. Kistler, and E. André, "The Social Signal Interpretation (SSI) Framework: Multimodal Signal Processing and Recognition in Real-time," in Proceedings of the 21st ACM International Conference on Multimedia, ser. MM '13. New York, NY, USA: ACM, 2013, pp. 831–834. [Online]. Available: http://doi.acm.org/10.1145/2502081.2502223

[14] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. Lim, and J. Suri, "Heart rate variability: a review," Medical and Biological Engineering and Computing, vol. 44, no. 12, 2006, pp. 1031–1051.

[15] J. Pan and W. Tompkins, "A real-time QRS detection algorithm," IEEE Transactions on Biomedical Engineering, 1985, pp. 230–236.

[16] R. Berger, S. Akselrod, D. Gordon, and R. Cohen, "An efficient algorithm for spectral analysis of heart rate variability," Biomedical Engineering, IEEE Transactions on, no. 9, 2007, pp. 900–904.

[17] W. Winton, L. Putnam, and R. Krauss, "Facial and autonomic manifestations of the dimensional structure of emotion," Journal of Experimental Social Psychology, vol. 20, no. 3, 1984, pp. 195–216.

[18] P. Lang, M. Greenwald, M. Bradley, and A. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," Psychophysiology, vol. 30, no. 3, 1993, pp. 261–273.

[19] P. Lang, "The emotion probe," American Psychologist, vol. 50, no. 5, 1995, pp. 372–385.

[20] D. Levinson and R. Edelberg, "Scoring criteria for response latency and habituation in electrodermal research: a critique," Psychophysiology, vol. 22, no. 4, 1985, pp. 417–426.

[21] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," Journal of neuroscience methods, vol. 190, no. 1, 2010, pp. 80–91.

[22] E. Garrett, "The bateman function revisited: a critical reevaluation of the quantitative expressions to characterize concentrations in the one compartment body model as a function of time with first-order invasion and first-order elimination," Journal of Pharmacokinetics and Pharmacodynamics, vol. 22, no. 2, 1994, pp. 103–128.

[23] A. Camm, M. Malik, J. Bigger, G. Breithardt, S. Cerutti, R. Cohen, P. Coumel, E. Fallen, H. Kennedy, R. Kleiger et al., "Heart rate variability: standards of measurement, physiological interpretation, and clinical use," Circulation, vol. 93, no. 5, 1996, pp. 1043–1065.

[24] M. Benedek and C. Kaernbach, "Decomposition of skin conductance data by means of nonnegative deconvolution," Psychophysiology, vol. 47, no. 4, 2010, pp. 647–658.

[25] S. Mitra, "Digital signal processing. a computer-based approach. 2001."

[26] F. Heijden, R. Duin, D. Ridder, and D. Tax, Classification, Parameter Estimation and State Estimation. John Wiley, 2004.

[27] R. Duda, P. Hart, and D. Stork, Pattern classification. 2nd edition, John Wiley and Sons, 2001.

[28] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," The Annals of Statistics, vol. 38, no. 2, 2000, p. 337 374.

[29] W. Hrdle and L. Simar, Applied multivariate statistical analysis. Springer, 2007.

[30] Affectiva Inc., "Liberate yourself from the lab : Q Sensor measures EDA in the wild (White Paper)," 2012.

Trace-based Task Tree Generation

Patrick Harms, Steffen Herbold, and Jens Grabowski
Institute of Computer Science
University of Göttingen
Göttingen, Germany
E-mail: {harms,herbold,grabowski}@cs.uni-goettingen.de

***Abstract*—Task trees are a well-known way for the manual modeling of user interactions. They provide an ideal basis for software analysis including usability evaluations if they are generated based on usage traces. In this paper, we present a method for the automated generation of task trees based on traces of user interactions. For this, we utilize usage monitors to record all events caused by users. These events are written into log files from which we generate task trees. We validate our method in three case studies.**

Keywords-task; tree; generation; usage-based; traces.

I. Introduction

Task trees are a well-known method to model user interactions as, e.g., done in [1]. They provide a structure to define how interactions are intended by the interaction designer [2]. They can also be used for comparing expected and effective user behavior as a basis for a semi-automatic usability evaluation [1]. Task trees are usually defined manually at design time [3]. For websites, they can also be generated based on existing Hyper-Text Markup Language (HTML) source code [4]. In both approaches, they do not describe effective user behavior but either expected or possible user behavior.

In this paper, we present an approach for automatically generating task trees based on recordings of user interactions. Such generated task trees represent the effective behavior of users and can, therefore, be used for usage analysis, e.g., in the context of usability evaluations. The results of a usage analysis can be used for optimizing software with respect to the user's needs. Throughout the remainder of this paper, we use the analysis of websites as a running example. However, our approach is designed for event-driven software in general including all kinds of desktop applications.

Task models are used to describe user actions. Task trees are one possible variant for modeling tasks. The concept of task trees is applied, e.g., in Goals, Operators, Methods, and Selection Rules (GOMS) [5], TaskMODL [6], and ConcurTaskTrees [7] [8]. We reuse the basic concept of task trees, but apply it in a simplified manner.

There have been several attempts to generate task trees automatically. For example, the Convenient, Rapid, Interactive Tool for Integrating Quick Usability Evaluations (CRITIQUE) [9] creates GOMS models based on recorded traces. A similar approach is proposed by John et al. [10]. ReverseAllUIs [4] generates task trees based on models of the Graphical User Interface (GUI). The resulting task trees represent all available interactions a user can perform. In contrast to our work, these approaches do not generate task trees that represent the effective behavior of the users, but only a simplified or complete task tree of a website.

A further attempt to identify reoccurring user behavior is programming by example. Here, user actions are recorded to determine reoccurring action sequences. The system then offers the user an automation of the identified action sequence. An example of this work can be found in [11]. These approaches only attempt to locally optimize the usability, whereas we adopt a global view on the system.

Generating task trees for user actions is similar to the inference of a grammar for a language. The user actions are the words of a language that the user "speaks" to the software. The task tree is the grammar defining the language structure. However, current approaches for grammatical inference require the identification of sentences of the language before the derivation of the grammar [12]. This is not feasible for our approach as the recorded user actions do not follow such a structure. For example, a user may interrupt a task execution, which would lead to an incomplete sentence.

The remainder of this paper is structured as follows: First, we introduce our approach and the respective terminology in Section II. Then, we describe an implementation in Section III and present three case studies in which we tested the feasibility of our approach in Section IV. We conclude with a discussion and an outlook on planned future work.

II. Trace-based Task Tree Generation

In this section, we introduce our process for generating task trees. We commence with the definition of terms that we use in this paper. Then, we describe our approach of tracing users of a website. Finally, we provide details about the generation of task trees based on the traces.

A. Terminology

Users utilize a website by performing elementary *action*s. An action is, e.g., clicking with the mouse on a button, typing some text into a text field, or scrolling a page. Actions cause *event*s to occur on a website, also known as Document Object Model (DOM) events. For example, clicking with a mouse causes an `onclick` event. Typing a text into a text field causes an `onchange` event on the text field. Events are a representation of actions. For each action there is a mapping to an event caused by performing the action.

To execute a specific *task* on a website, a user has to perform several actions. For example, for logging in on a website, a user must type in a user name and a password into two separate text fields and click on a confirmation button.

 ISBN: 978-1-61208-325-4

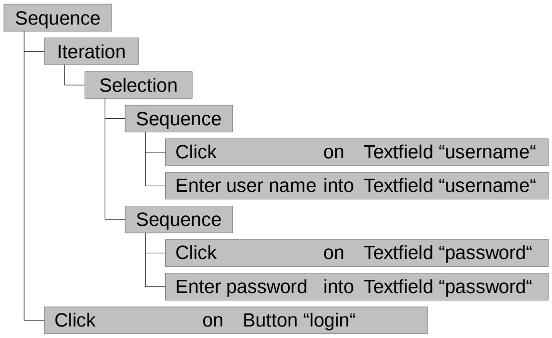


Figure 1. Example for a task tree

Tasks and actions can be combined to form higher level tasks. For example, the task of submitting an entry on a forum website comprises a *subtask* for logging in on the website as well as several actions for writing the forum entry and submitting it. Therefore, tasks and actions form a tree structure called a *task tree*. The leaf nodes of a task tree are the actions a user must perform to fulfill the overall task. The overall task itself is the root node of the task tree. The intermediate nodes in the task tree structure the overall task into subtasks.

A task defines a *temporal relationship* for its children, which specifies the order in which the children (subtasks and actions) must be executed to fulfill the task. Different task modeling approaches use different temporal relationships [8]. In our work, we consider the temporal relationships *sequence*, *iteration*, and *selection*. If a task is a sequence, its children are executed in a specified order. If a task is an iteration, it has only one child, which can be executed zero or more times. If a task is a selection, only one of its children is performed. A leaf node in a task tree has no children and does, therefore, not define a temporal relationship.

An example for a task tree is shown in Figure 1. It represents the actions to be taken to perform a login on a website. The actions are the leaf nodes. The temporal relationships of their parent nodes define the order in which the actions have to be performed. The task starts with an iteration of a selection. The possible variants are entering a user name or a password in the respective fields. The user may enter and change his user name and password several times. The overall task is completed after the user clicks the login button.

B. User Interaction Tracing

The first step in our approach is tracing user actions on a website. This is done by recording the events caused by the actions of a user. We achieve this by integrating a monitoring module in the website. This module is invisible to the user and has minimal effect on the implementation, performance, and stability of the website [13]. The resulting sequence of events is encrypted, sent to a server, and stored in a log file. A recorded sequence of events is called a trace.

A simplified example of a trace is shown in Figure 2. It lists the events recorded for a login of a user on a website. The login comprises the entering of the user name and the password in the respective text fields, as well as a confirmation by clicking on the login button. As the user initially entered a wrong user name, he reenters it a second time.

1.	Left mouse button click	on	Textfield with id username
2.	Text input „usr“	on	Textfield with id username
3.	Left mouse button click	on	Textfield with id username
4.	Text input „user“	on	Textfield with id username
5.	Left mouse button click	on	Textfield with id password
6.	Text input „“	on	Textfield with id password
7.	Left mouse button click	on	Button with name „login“

Figure 2. Example for a trace

Conceptual Design	Types of entities and their relationships
Semantic Design	Functions to modify entities
Syntactical Design	Steps to take for executing functions on entities
Lexical Design	Physical execution of steps to execute functions on entities

Figure 3. Levels of design

C. Task Tree Generation

To describe the process for generating task trees based on traces, we introduce the levels of design, which are important for structuring task trees. We then describe the creation of the initial task tree, which is afterwards refined and condensed using temporal relationships.

1) Basic Approach: When designing GUIs, four levels of design are considered: conceptual design, semantic design, syntactical design, and lexical design. They are shown in Figure 3. The conceptual design describes the types of entities that are to be edited with a software [14], as well as their relationships [15]. For example, in a system for managing addresses, addresses and persons are the entity types. These entity types are related, because a person may be assigned zero or more addresses.

The semantic design specifies functions to edit the entities defined in the conceptual design [14]. For the address management example, this includes adding, editing, and deleting addresses and persons. The syntactical design specifies the steps to execute a function defined in the semantic design [14]. For example, adding a new address is comprised of steps like adding a street name, a city, and a zip code. At the most detailed level, the lexical design specifies means of physically performing steps defined in the syntactical design [14]. In the example, defining a street of an address includes clicking on the respective text field and typing the street name.

In our approach, we map the semantic, syntactical, and lexical levels of design onto task trees. For each function specified in the semantic design, there exists a task for executing that function. Hence, there is one task tree for each function in the semantic design. The syntactic design is a decomposition of functions into individual steps for function execution. This decomposition corresponds to the definition of subtasks and their temporal relationships within task trees. The actions on the lexical level of design are represented through the leaf nodes of task trees. As we record the events mapped to the respective actions, we refer to the leaf nodes as event tasks. Event tasks are considered normal tasks with the constraint of not having children and not defining a temporal relationship.

Using this basic approach, we create task trees starting from the leaf nodes, i.e., from the event tasks. For each event in a trace, we generate an event task. All event tasks are stored in an ordered list in the order the respective events were recorded.

 ISBN: 978-1-61208-325-4

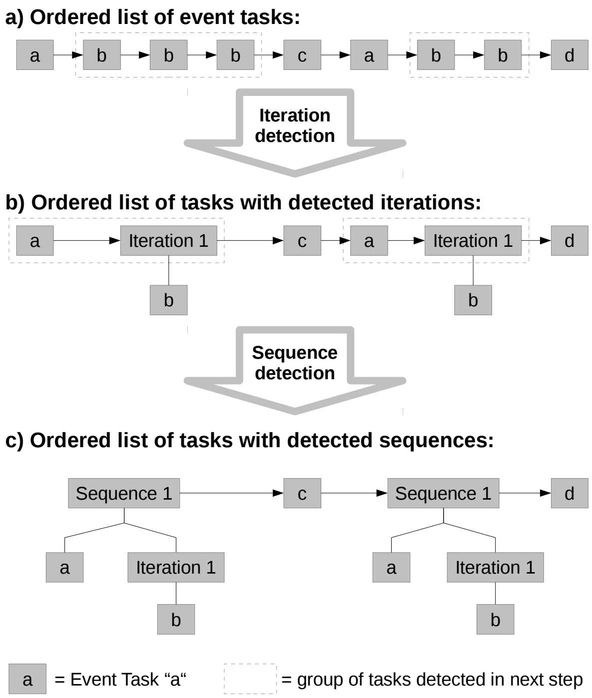


Figure 4. Example for the detection of iterations and sequences

An example is shown in Figure 4a where each grey rectangle denotes an event task and the arrows denote their order.

2) Iteration Detection: The ordered list of event tasks may contain identical tasks that occur subsequently. For example, the user might have clicked several times on the same button. Such tasks are represented in task trees as iterations. Therefore, we scan the list of event tasks for iterations of identical tasks. If we observe an iteration, we generate a new task node of type iteration. This node gets the iterated event task as its single child. We then replace each occurrence of an iteration of the event task in the ordered list with the new iteration task node. Several subsequently occurring identical event tasks are herewith replaced by a single task node of type iteration. An example for this approach is shown in Figure 4. There, Event Task *b* is iterated several times (denoted by dotted boxes in Figure 4a). We replace these occurrences in the task list with single iteration nodes (Figure 4b).

3) Sequence Detection: After the iteration detection, we scan the list of tasks for identical subsequences. For the subsequence occurring most often and which is, therefore, most likely an occurrence of a logical subtask, we generate a new task node of type sequence. Its children are the tasks belonging to the subsequence. Each occurrence of the identified subsequence in the task list is replaced with the new sequence task node. An example is shown in Figure 4. There, the subsequence of Event Task *a* and Iteration *1* occurs most often (two times) and is, therefore, replaced through task nodes representing this sequence.

The subsequences replaced through the sequence detection can have any length. At the minimum, they have a length of two. Our algorithm searches for the longest subsequences occurring most often and replaces it accordingly. If several subsequences have the same maximum occurrence count, we replace only the longest one. If several subsequences have the same maximum count and the same maximum length, we replace only the subsequence occurring first in the ordered list.

4) Repetition of Detections: The iteration and sequence detection on the list of tasks are repeated alternately until no more replacements are done. Each time an iteration detection is done, all iterations are detected and replaced. This also includes iterations of detected sequences. For each sequence detection the longest sequence occurring most often is replaced. A detected sequence may include already detected sequences and iterations. For example, in Figure 4c the detected sequence contains a previously detected iteration.

If no more iterations or sequences are detected, the algorithm stops. The resulting task list contains detected task trees as well as event tasks, which were neither iterated nor part of a sequence occurring more than once. The detected task trees represent the lexical, syntactical and semantic level of design. The more recorded events are processed, the more complex and deeper task trees are created.

Within a recording of only one user session, specific subsequences occur only once. An example is the login process, which is usually done only at the beginning of a recorded user session. With our approach, such regularly occurring subsequences would not be detected if only one session was considered. Therefore, we consider several sessions of different users at once for counting the number of occurrences of subsequences. Due to this, we also detect subsequences occurring seldom in individual sessions but often with respect to all recorded users of the website.

D. Usability Evaluation

We utilize the generated task trees for automated usability evaluations. For this, we consider violations of generally accepted usability heuristics (e.g., as provided in [16]) and define patterns for their reflection in task trees. We then filter our task trees for these patterns and reason on potential usability defects. This is possible, as the generated task trees represent effective user behavior. However, this work is still in its infancy and, therefore, not described in more detail.

III. Proof-of-concept Implementation

To show that our method is feasible, we implemented it based on the tool suite for Automatic Quality Engineering of Event-driven Software (AutoQUEST) [17]. The AutoQUEST platform provides diverse methods for assessing the quality of software. AutoQUEST's internal algorithms operate on abstract events, which makes AutoQUEST independent of the platform of an assessed software. AutoQUEST's modular architecture allows the extension with modules to support algorithms for quality assurance, as well as feeding AutoQUEST with events of an yet unsupported software platform. In the following, we describe how we utilized and extended AutoQUEST to implement our method.

A. User Interaction Tracing

AutoQUEST provides basic functionality for tracing user actions. For this, it uses techniques from GUI testing. A popular approach for GUI testing is capture/replay [18], a technique where the tester interacts with the software and a capture tool records the executed actions. Afterwards, the actions are automatically executable with a replay tool in order to generate automated software tests. AutoQUEST uses the capturing to trace users and we reused these capabilities for our implementation for tracing users of websites.

 ISBN: 978-1-61208-325-4

```
<event type="onclick">
 <param name="X" value="87"/>
 <param name="Y" value="213"/>
 <param name="target" value="id1"/>
 <param name="timestamp" value="1375177632056"/>
</event>
<event type="onscroll">
 <param name="scrollX" value="-1"/>
 <param name="scrollY" value="-1"/>
 <param name="target" value="id2"/>
 <param name="timestamp" value="1375177632900"/>
</event>
```

Figure 5. Example for a trace recorded with AutoQUEST's HTML monitor

AutoQUEST provides several platform specific plug-ins to trace the usage of software. This includes plug-ins for the Microsoft Foundation Classes, Java Foundation Classes, and websites. All plug-ins are comprised of a monitor to trace software usage and a trace parser to feed the recorded events into AutoQUEST. The monitors can be integrated with minimal effort into the software to be monitored. For example, for monitoring a website only a JavaScript needs to be added to each of the pages of the website. In modern content management systems, this can be configured centrally and easily. The JavaScript is served by a monitoring server provided with AutoQUEST. After the integration of the JavaScript in the website, it automatically records events caused by user actions. After a specific amount of events is recorded, or if the user switches the page, the script sends the events to the AutoQUEST server which stores them into log files.

An excerpt of a trace of AutoQUESTs website monitor showing a mouse click and a scroll event on a web page is shown in Figure 5. Both events denote their respective type, a timestamp, and meta information like the coordinates in the click event. Furthermore, both events refer to a target, i.e., the element of the webpage, on which the event was observed. The identifiers of the targets can be resolved through other information stored in the log file, as well.

B. Task Tree Generation

For our proof of concept, we extended AutoQUEST with capabilities to generate task trees based on traces. The implementation follows the overall process described in Section II-C. The implementation of the iteration detection is straightforward and, therefore, not described in more detail.

1) Sequence Detection Implementation: For identifying and counting subsequences occurring several times, we reused and extended a data structure provided with AutoQUEST called trie [13]. A trie in AutoQUEST is a tree structure used for representing occurrences of subsequences in a sequence. In our case, we use the trie for representing subsequences of tasks in the ordered list of tasks considered for the next sequence detection. An example for a trie is shown in Figure 6.

Each node in a trie represents a task subsequence. The length of the represented subsequence is equal to the distance of the node to the root node of the trie. The root node of the trie represents the empty subsequence. The children of the root node (in Figure 6 all nodes on Level 1) represent the subsequences of length 1 occurring in the trace, i.e., all different tasks. The grand children of the root node (in Figure 6 all nodes on Level 2) represent the subsequences of length two as their distance to the root node is two, etc. The subsequence represented by a node can be determined by following the path through the trie starting from the root node and ending at the respective node. The length of the longest subsequence represented through a node in the trie is defined as the depth of the trie. The depth of the trie in Figure 6 is three.

Each node in a trie is assigned a counter. This counter defines the number of occurrences of the subsequence represented by the node. The counter of the root node is ignored. The example trie in Figure 6 represents the event tasks for the trace of Figure 2. The trie shows that the event task of clicking on the user name text field occurs twice and that both times it is succeeded by entering some text, i.e., a user name, into the text field. The event of clicking the login button is not succeeded by any other event task.

We calculate a trie each time a sequence detection on the ordered list of tasks is done. Based on the trie, we are able to identify the longest subsequence of tasks with a minimal length of two occurring most often. The number of occurrences is determined through the counts assigned to each node in the trie. The length of the subsequence is determined by the distance of the trie node representing the most occurring subsequence to the root node of the trie.

If the length of the identified subsequence is identical to the depth of the trie, we cannot decide if there is a longer subsequence with the same count. We, therefore, increase the depth of the trie until the depth is larger than the length of the longest subsequence occurring most often. In Figure 6, the longest subsequence occurring most often is clicking on the user name text field and entering a user name. This subsequence occurs twice and there is no other subsequence of the same or a longer length occurring more often. Therefore, all occurrences of this subsequence in the ordered list of tasks is replaced through a task node of type sequence.

2) Comparison of tasks: An important challenge in our implementation was the comparison of tasks. Tasks need to be compared very often either for compiling the trie or for detecting iterations. For an effective task generation, some tasks must be considered equal although they are different. An example is a task and an iteration of this task. Both must be considered identical if the iteration is executed only once. Another example is shown in Figure 6. The represented trie contains nodes for the event tasks representing the entering of text into the user name text field. Although different text is entered in the respective events, the respective event tasks need to be considered identical for a correct trie calculation. Therefore, we implemented a mechanism to be able to perform complex task comparisons. In addition to other comparisons, it is able to compare a task A with an iteration of a task B and considers them as equal if task A is equal to task B.

IV. Case Studies

For the validation of our approach, we performed three case studies. For the first case study, we traced the interaction of users of our research website [19]. We integrated the HTML monitor of AutoQUEST in our content management system. We then recorded interactions of more than 700 users over a period of 6 months. Afterwards, we fed the gathered traces containing more than 25,000 events into AutoQUEST and generated over 600 task trees based on this. This case study showed that the task tree generation was feasible in general. The generated task trees represented user behavior occurring several times. As an example, several users opened the initial web page and navigated to our teaching page. From there, they navigated to the information about a specific lecture.

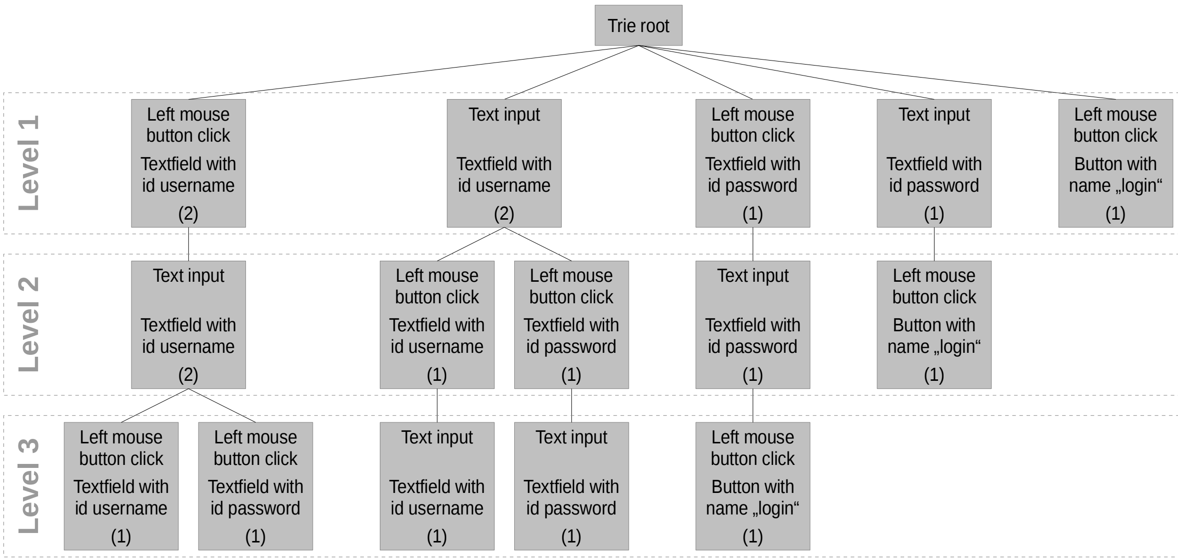


Figure 6. Trie generated based on the trace in Figure 2

The first case study also revealed that our mechanism must be careful with respect to privacy protection. Our research website includes a log-in mechanism for being able to change its content. The first version of the tracing mechanism also traced user names and passwords of all users that logged in on the website. As this was a severe security issue, we adapted the tracing mechanism to ignore password fields in general. Furthermore, a website can be instrumented in a way, so that contents of selected text fields, e.g., fields for entering a user name, are not traced anymore.

In our second case study, we traced the users of an application portal of our university over a period of 3 months. This case study traced over 500 users producing more than 150,000 events resulting in 5,320 generated task trees. When feeding this data into AutoQUEST, we initially observed performance problems of our approach. Especially, the large number of distinct events caused the creation of a large trie for sequence detection. We, therefore, implemented several optimizations. For example, click events on the same button but with different coordinates are now treated as the same event task. However, click events on other website elements are still considered different, if their coordinates differ.

The second case study showed that our approach is able to correctly identify effective user behavior. The application portal also provides a login mechanism. Our task tree generation created several different task trees for the login process of users. One of them showed the behavior of those users using the mouse to set the focus on the password field after having entered the user name. The other task trees showed the usage of the tabulator key instead. A visualization of the second login variant as displayed by AutoQUEST is shown in Figure 7. This presentation is a further extension done for AutoQUEST in the context of our work. The example shows, that many iterations are generated in the task tree. This is due to the fact, that some users corrected the entered data several times. Furthermore, if the users entered wrong credentials, the website returned to the same view and the users started the login process again.

In our third case study, we developed a sandbox example to validate the task tree generation with a subsequent usability

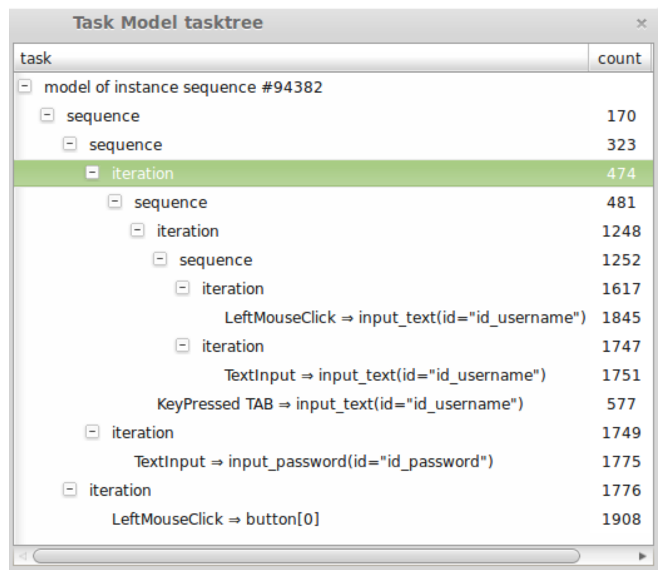


Figure 7. Task tree generated in the context of the second case study

evaluation. The sandbox contains several distinct views, each focused on a specific usability defect. We recorded our interactions on the sandbox, generated task trees, and performed an automated usability evaluation. The case study showed, that an automated usability evaluation is possible in general. However, the current implementation produces many false positives and is not mature enough to be described in more detail.

V. Discussion of our Method

The generated task trees represent the effective user behavior. This is important to analyze the usage of a monitored website, e.g., with respect to usability. Currently, our approach is not able to identify distinct ways of executing semantically equal tasks. As an example, the different ways of filling out the login form in the case studies are treated as different tasks in the generated task tree.

At each repetition, the detection of subsequences chooses the longest sequence occurring most often and replaces it as described. This heuristic prefers shorter sequences as the count decreases with an increasing sequence length. The resulting task trees are, therefore, deeply structured. Hence, it would be better to apply a more sophisticated heuristic such as selecting a subsequence occurring more seldom but being much longer.

VI. Summary and Outlook

In this paper, we described a method for generating task trees based on tracing user interactions. We implemented this method for websites and performed three case studies to validate its feasibility.

In our future work, we will improve and extend the task tree generation. We especially focus on the detection of an enhanced set of temporal relationships not considered in our work, yet. An example is the detection of selections of different approaches for executing the same task. We also plan to support a manual merging of such tasks to be able to treat them as identical in a subsequent usage analysis. Furthermore, we plan to implement both, a better heuristic for detecting more intuitive subsequences, as well as a flattening algorithm for reducing the complexity of the generated task trees. In addition, we improve the existing AutoQUEST plug-ins and implement plug-ins for further platforms, e.g., for operating systems with a focus on touch-based interaction. Finally, we improve the automated usability evaluation based on the generated task trees.

Acknowledgment

This work was done in the context of the project MIDAS (Model and Inference Driven - Automated testing of Services Architectures).

References

[1] F. Paternò, "Tools for remote web usability evaluation," in HCI International 2003. Proceedings of the 10th International Conference on Human-Computer Interaction. Vol.1, vol. 1. Erlbaum, 2003, pp. 828–832.

[2] L. Paganelli and F. Paternò, "Tools for remote usability evaluation of web applications through browser logs and task models," Behavior Research Methods, vol. 35, 2003, pp. 369–378.

[3] F. Paternò, "Model-based tools for pervasive usability." Interacting with Computers, vol. 17, no. 3, 2005, pp. 291–315.

[4] R. Bandelloni, F. Paternò, and C. Santoro, "Engineering interactive systems," J. Gulliksen, M. B. Harning, P. Palanque, G. C. Veer, and J. Wesson, Eds. Berlin, Heidelberg: Springer-Verlag, 2008, ch. Reverse Engineering Cross-Modal User Interfaces for Ubiquitous Environments, pp. 285–302.

[5] Q. Limbourg and J. Vanderdonckt, "Comparing task models for user interface design," in The Handbook of Task Analysis for Human-Computer Interaction, D. Diaper and N. Stanton, Eds. Mahwah: Lawrence Erlbaum Associates, 2004.

[6] H. Trætteberg, Model-based user interface design. Information Systems Group, Department of Computer and Information Sciences, Faculty of Information Technology, Mathematics and Electrical Engineering, Norwegian University of Science and Technology, May 2002.

[7] F. Paternò, C. Mancini, and S. Meniconi, "ConcurTaskTrees: A diagrammatic notation for specifying task models," in Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction, ser. INTERACT '97. London, UK, UK: Chapman & Hall, Ltd., 1997, pp. 362–369.

[8] F. Paternò, "ConcurTaskTrees : An engineered approach to model-based design of interactive systems," The Handbook of Analysis for HumanComputer Interaction, 1999, pp. 1–18.

[9] S. E. Hudson, B. E. John, K. Knudsen, and M. D. Byrne, "A tool for creating predictive performance models from user interface demonstrations," in Proceedings of the 12th annual ACM symposium on User interface software and technology, ser. UIST '99. New York, NY, USA: ACM, 1999, pp. 93–102.

[10] B. E. John, K. Prevas, D. D. Salvucci, and K. Koedinger, "Predictive human performance modeling made easy," in Proceedings of the SIGCHI conference on Human factors in computing systems, ser. CHI '04. New York, NY, USA: ACM, 2004, pp. 455–462.

[11] A. Cypher, "Eager: programming repetitive tasks by example," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '91. New York, NY, USA: ACM, 1991, pp. 33–39.

[12] A. D'Ulizia, F. Ferri, and P. Grifoni, "A survey of grammatical inference methods for natural language learning," Artif. Intell. Rev., vol. 36, no. 1, Jun. 2011, pp. 1–27.

[13] S. Herbold, "Usage-based Testing of Event-driven Software," Ph.D. dissertation, University Göttingen, June 2012 (electronically published on http://webdoc.sub.gwdg.de/diss/2012/herbold/ [retrieved: 1, 2014]), 2012.

[14] R. J. Jacob, "User interface," in Encyclopedia of Computer Science, ser. Encyclopedia of Computer Science, A. Ralston, E. Reilly, and D. Hemmendinger, Eds. Nature Publishing Group London, 2000, pp. 1821–1826.

[15] J. Foley, Computer Graphics: Principles and Practice, ser. Systems Programming Series. Addison-Wesley, 1996.

[16] U.S. Department of Health & Human Services. Usability.gov - improving the user experience - guidelines. [Online]. Available: http://guidelines.usability.gov/ [retrieved: 1, 2014] (2013)

[17] S. Herbold and P. Harms, "AutoQUEST - Automated Quality Engineering of Event-driven Software," submitted to the Testing Tools Track of the International Conference on Software Testing (ICST) 2013, unpublished.

[18] J. H. Hicinbothom and W. W. Zachary, "A Tool for Automatically Generating Transcripts of Human-Computer Interaction," in Human Factors and Ergonomics Society 37th Annual Meeting, vol. 2 of Special Sessions, 1993, p. 1042.

[19] Software Engineering for Distributed Systems Group. Software Engineering for Distributed Systems. [Online]. Available: http://www.swe.informatik.uni-goettingen.de/ [retrieved: 1, 2014] (2014)

 ISBN: 978-1-61208-325-4

Modeling User's State During Dialog Turn Using HMM For Multi-modal Spoken Dialog System

Yuya Chiba, Akinori Ito
Graduate School of Engineering,
Tohoku University
Email: {yuya, aito}@spcom.ecei.tohoku.ac.jp

Masashi Ito
Department of Electronics and Intelligent System,
Tohoku Institute of Technology
Email: itojin@totec.ac.jp

***Abstract*—Conventional spoken dialog systems cannot estimate the user's state while waiting for an input from the user because the estimation process is triggered by observing the user's utterance. This is a problem when, for some reason, the user cannot make an input utterance in response to the system's prompt. To help these users before they give up, the system should handle the requests expressed by them unconsciously. Based on this assumption, we have examined a method to estimate the state of a user before making an utterance by using the non-verbal behavior of the user. The present paper proposes an automatic discrimination method by using time sequential non-verbal information of the user. In this method, the user's internal state is estimated using multi-modal information such as speech, facial expression and gaze, modeled using a Hidden Markov Model (HMM).**

Keywords- multi-modal information processing; user's state; spoken dialog system

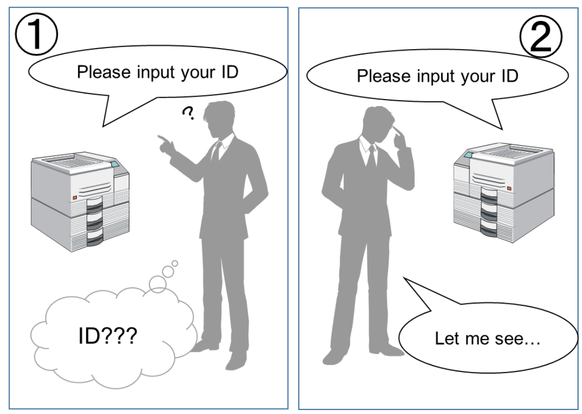


Figure 1. Target user's state

I. Introduction

Most spoken dialog systems estimate the user's internal state to generate an appropriate response to the user. Many researches on user modeling have been conducted such as emotion [1], [2], preference [3] and familiarity with the system [4]. These methods implicitly assume that the user always gives some responses to the system's prompt. However, not all users can use the system proficiently. For instance, a user may abandon a session without uttering a word if he or she cannot understand the meaning of the system's prompt, or could take a long time to consider how to answer the prompt. The user model which does not depend on linguistic information is needed to help these users appropriately. To tackle this problem, we have assumed two basic internal states of a user who cannot make an utterance. The first one is the state where the user does not know what to input, and the second one is when the user is considering how to answer the system's prompt. We call these states "state A" and "state B", respectively. Although our definitive goal is building a spoken dialog system that can help the user in an optimum manner, we focus on the estimation method of user's state in this report. Since discrimination must be processed after the user gets the floor and before he/she makes the input utterance to the system, we denote them as the user's internal state "during a dialog turn." Figure 1 shows typical examples of these user's states.

In a human-human dialog, interlocutors converse while more or less estimating the internal state of the dialog partner based on the feeling that the other person knows the answer to the question (in other words, whether other interlocutors could respond to his/her utterance or not). This ability is called "Feeling of Another's Knowing" (FOAK) [5][6]. It is thought that introducing such manner of human dialog into a dialog system would improve its performance.

To estimate the user's state during a dialog turn, we used visual and acoustic features instead of linguistic contents of the utterance. Bi-modal feature fusion has been examined in the field of emotion recognition [7][8]. In the present research, almost the same features as these studies were employed because the user's state during a dialog turn has similar aspects to emotion. Besides, successive estimation is required to achieve the goal of our work because there is no explicit trigger for processing the user's utterance.

In the previous report [9], we used a multi-stage neural network to integrate multi-modal features and estimate the user's state frame by frame. However, this method has the problem that it cannot capture the temporal variation of the features. Based on this result, the present study examines an estimation method which uses multi-stream Hidden Markov Model (HMM) to model the local temporal variation of the audio-visual feature sequence. In this method, the likelihoods of the user's states are obtained continuously. The performance of the proposed method is evaluated by discrimination examination.

This paper is organized as follows. Collection of the experimental data is described in Section 2. Then estimation method

 ISBN: 978-1-61208-325-4

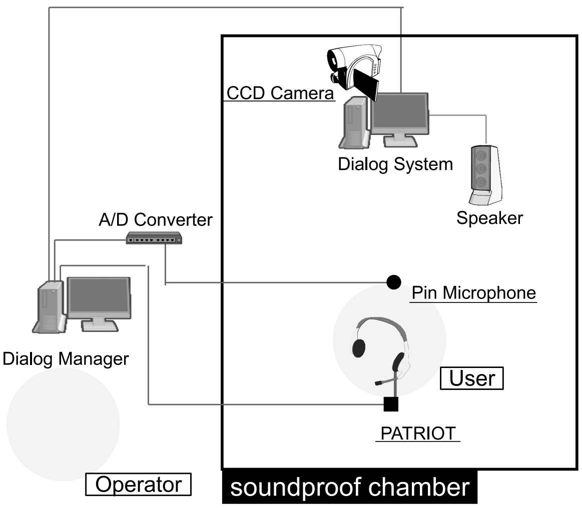


Figure 2. Experimental circumstance

TABLE I. EVALUATION RESULTS

State A	State B	Neutral	Total
59	195	538	792

using multi-stream HMM is introduced in Section 3. The multi-modal features employed in this paper is described in Section 4. Finally, the results of the experiments are presented in Section 5.

II. EXPERIMENTAL DATA

Experimental data were collected on the Wizard of Oz basis. The dialog experiments were conducted in a soundproof chamber. We implemented a question-and-answer task in which the system posed questions and the subjects answered them. The task was designed to make the user as embarrassed as possible. The questions were about common knowledge or a number memorized in advance, such as "Please input current date." and "Please input your ID number." Additionally, an agent with a simple cartoon-like face was projected on the monitor to keep the subjects' attention. Figure 2 shows an experimental circumstance. We employed 16 subjects (14 males and 2 females). The subjects wore a lapel microphone. To record an image of the subjects' frontal face, a CCD camera was installed above the monitor in front of the subjects. The operator remained outside of the chamber and controlled the agent remotely. The audio signal was recorded in PCM format at 16 kHz sampling, 16-bit quantization. The recorded video clips were stored as AVI files with 24-bit color depth, 30 frame/s. After the experiment, we separated the dialog into sessions; one session included one interchange of the system's prompt and the user's response. Here, we defined the length of the segment between the end of the system's prompt and the beginning of the user's input utterance as "latency". Sessions with more than 5.0 s latency were labeled by five evaluators.

Table I shows the results of the evaluation. The label of each session was chosen by majority vote of evaluators (Fleiss' $\kappa = 0.22$). One session was excluded because the acoustic feature could not be extracted due to overlapping utterances.

III. DISCRIMINATION METHOD USING MULTI-STREAM HMM

The spoken dialog system has to detect whether the user needs help or not as soon as possible, because the ultimate purpose of our work is to build a system that responds to a user who has difficulty in answering a question. Therefore, we need incremental evaluation of the user's internal state, and the system should help the user just after detecting the stagnation of the dialog. We therefore observed the time sequence of the features of the user, and fed the features to the classifier frame by frame. We assume the user's non-verbal behavior was recorded continuously during the dialog with the microphone and the CCD camera. In the previous paper, we have examined the method to estimate the user's state by using a single audio-visual feature frame as a feature vector of the classification. However, capturing the temporal variation of the feature sequence is considered to be essential for the estimation. For instance, the user thinking of the response to the system (i.e., state B) tends to emit long fillers, and the user cannot understand the meaning of the prompt (i.e., state A) might move his/her eyes frequently. Therefore, we proposed the method that extracts the segment of the feature frames as a vector and feeds the feature vector to the HMM in order to represent temporal characteristics of the feature sequence. Additionally, we used a multi-stream HMM as the classifier. The multi-stream HMM can fit the distribution of the multi-modal features efficiently by dealing with the feature sequence belonging to the different modality as the different stream. The score of the user's state is emitted frame by frame, and they are integrated at the final stage to decide the discrimination result. The topology of the HMM was ergodic; therefore, the HMM has transitions between all states.

The output probability of the state j, $b_j(\boldsymbol{o}_t)$ is denoted as follows:

$$b_j(\boldsymbol{o}_t) = \prod_{s=1}^{S} \left[\sum_{m}^{M_s} c_{jsm} \mathcal{N}(\boldsymbol{o}_{st}; \boldsymbol{\mu}_{jsm}, \boldsymbol{\Sigma}_{jsm}) \right]^{\lambda_s} \quad (1)$$

where, $\boldsymbol{o}_{st}$ is the feature sequence belonging to stream s at time t, c_{jsm}, μ_{jsm}, Σ_{jsm} is the parameter of the output probability density function of the state j and represents weights of mixture, mean vector, and covariance matrix, respectively. Each stream corresponds to each modality (indicated in the next section). $\mathcal{N}(\boldsymbol{o}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate Gaussian function, that is;

$$\mathcal{N}(\boldsymbol{o}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(\boldsymbol{o} - \boldsymbol{\mu})^t \Sigma^{-1} (\boldsymbol{o} - \boldsymbol{\mu})\} \quad (2)$$

These equations show that the multi-stream HMM emits the output log likelihood as the weighted sum of the output log likelihoods of each stream.

In addition, we show the overview of the construction of the segmental feature in Figure 3. Segmental feature f_t is constructed to have the past n frames from time t and obtained by shifting one frame at a time. Therefore, the segmental feature enables both the investigation of temporal characteristic of the feature sequence of the short segment and frame by frame estimation of the user's state. As shown in Figure 3 the number of frame n was set to 100, which is equivalent to one second in real time.

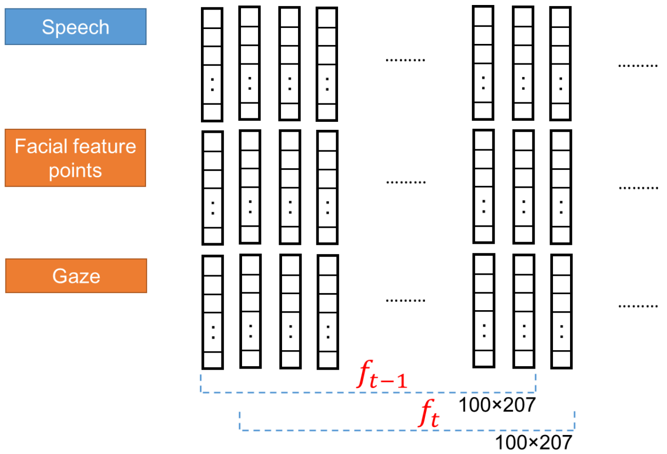


Figure 3. Segmental feature construction

TABLE II. CONDITIONS OF ACOUSTIC FEATURE EXTRACTION

	MFCC	ΔPitch	Zero cross ratio
Frame width	25.0 ms	7.5 ms	10.0 ms
Frame shift	10.0 ms	10.0 ms	10.0 ms

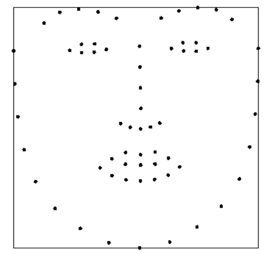


Figure 4. Model of facial feature points

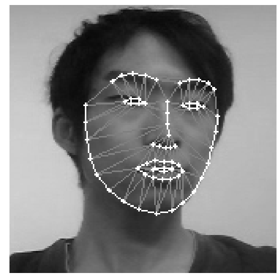


Figure 5. Result of feature extraction

IV. MULTI-MODAL SEQUENTIAL FEATURES

The target user's states are assumed to have similar aspects to emotion, and therefore we employed almost the same features used in the area of emotion recognition, such as the spectral features of the speech, intonation, zero cross ratio, facial feature points and gaze direction. In particular, it is suggested that emotion has a multi-modality nature [10], and most researches reported that recognition accuracy is improved by combining multi-modal information [7], [8].

A. Acoustic feature

To represent spectral characteristics, Mel-Frequency Cepstrum Coefficients (MFCC) is employed as a low-level acoustic feature. In our method, velocity and acceleration coefficients of MFCC including the log power were used. The total number of dimensions of MFCC is 39 and the frame length of calculating the time difference components is 5. Intonation of speech is represented by an $F0$ contour. The $F0$ was extracted by the normalized cross correlation, then converted to the log-scale. Since the $F0$ has large variation from speaker to speaker, a differential coefficient is used as the acoustic feature. In addition to the features mentioned above, the zero cross ratio is used to distinguish voiced and unvoiced segments. The basic conditions for extracting each acoustic feature are shown in Table II.

B. Facial feature points

Facial activity of the user is the most important feature among the visual information. To represent the facial activity, feature points of the face were extracted by Constraint Local Model (CLM) [11]. In this method, a model of the feature points is fitted after detecting the facial region from the whole image in the frame. Figure 4 shows a model of feature points and Figure 5 is an example of the result of fitting. The fitting error is mainly caused by misdetection of the facial region and occlusion. Although misdetection of the facial region was corrected by hand for the examination, the error caused by occlusion was not considered in the feature extraction; 5% of all sessions contained severe fitting errors. We used the relative coordinates of the feature points as the visual features. The number of feature points was 66 and the number of dimensions of the features was 132. The locations of feature points were normalized by the size of the facial region.

C. Gaze feature

Previous analysis showed the user's gaze action affects the evaluation of the user's state. Therefore, we used brightness feature of the eye region to represent the broad location of the user's iris indirectly. We employed Haar-like feature which has the fast calculation algorithm using the integral image. The Haar-like feature is extracted by applying filters depicted as Figure 6 to the image and originally used for object detection. We calculated the Haar-like feature from both eyes region obtained by CLM. As the Haar-like feature vector has high dimensionality, the principal component analysis (PCA) is applied for reducing the dimensionality of the feature vectors. After reducing the dimensions, the gaze feature has 34 dimension and cumulative contribution rate was about 95 %.

D. Feature synchronization

Finally, these features were synchronized because audio and visual information were extracted by different sampling rate. We synchronized the features by copying the previous visual feature values for each 10 ms. Therefore, the number of dimensions of a combined feature frame was always 207.

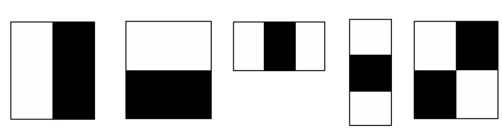


Figure 6. Haar-like filter

 ISBN: 978-1-61208-325-4

TABLE III. EXPERIMENTAL CONDITION

Experimental data	State A(59), State B(195)
Number of states	3, 4, 5
Number of mixture components of each stream	4

V. DISCRIMINATION EXAMINATION

A. Experimental condition

To evaluate the proposed method, we conducted a discrimination examination. Table. III shows the experimental condition. The number in parentheses indicates the number of sessions.

The previous work [9] showed that discrimination between the neutral state and the other states can be done by using the latency. Therefore, we focus on the discrimination between the state A and state B in the following experiment. Although the weights of streams should be decided to optimize the discrimination results, we fixed all stream weights to 1.0 to verify the effectiveness of estimation using HMM. The optimization method of the stream weights is a future work. In this paper, we changed the number of states of the HMM and evaluated the discrimination accuracy.

The definitive discrimination result was decided by comparing the average scores of each states. That is:

$$\hat{c} = \arg\max_{c} \frac{1}{T} \sum_{t=1}^{T} p_{tc} \tag{3}$$

where T is the length of the segment for which the state is estimated. In this experiment, we used the duration of each session as T. Here, the total accuracy tends to increase as the determined class leans toward state B because the amount of data is not uniformly distributed (see Table. I and Table. III); therefore, the harmonic mean (denoted as $Harm.$) was employed for measuring the performance. This is calculated by

$$Harm. = \frac{2 \cdot Corr_A \cdot Corr_B}{Corr_A + Corr_B} \times 100.0 \quad (\%) \tag{4}$$

where $Corr_A$ and $Corr_B$ represent the discrimination accuracy of state A and state B, respectively. The experiments were conducted by 5-fold cross validation.

B. Experimental result

Figure 7 shows the experimental results. We showed the result of the previous experiment [9] (Baseline in Figure 7) for comparison with the proposed method. The best performance was obtained when the number of states of HMM was 4 ($Harm. = 64.4\%$). This result was about 2 points higher than the baseline method, which used a neural network and a single feature frame. However, the improvement of the performance was not as large as expected, considering the additional computation cost. One of the reasons is the output distributions of each stream are not learned enough due to the shortage of the training data. The multi-stream HMM has an advantage in that it can control the importance of the output probability of each stream by changing the stream weights. Therefore, we will examine the method to optimize the stream weights to improve the performance in a future work.

VI. CONCLUSION

We examined the user modeling of dialog turn using a multi-stream HMM and segmental feature to represent temporal variation of an audio-visual feature sequence. The proposed

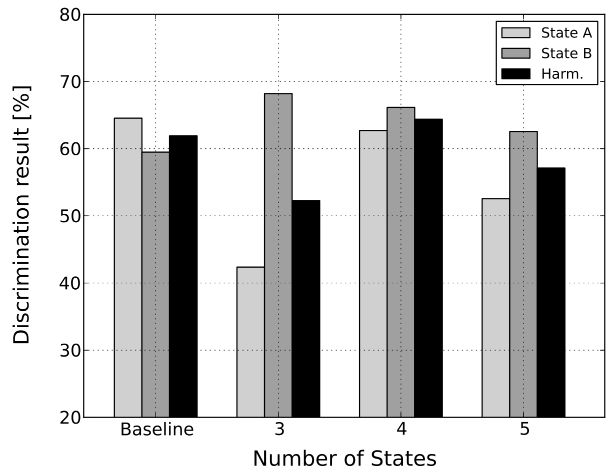


Figure 7. Discrimination results

method obtained the best result when the number of state was 4, and the result surpassed our previous work. On the other hand, we also observed the limitation of the discrimination performance of multi-stream HMM without optimizing the stream weights. In a future work, we will examine the method to decide the stream weights to improve the discrimination result.

REFERENCES

[1] K. Forbes-Riley and D. Litman, "Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor," Speech Communication, vol. 53, 2011, pp. 1115–1136.

[2] A. Metallinou *et al.* "Context-sensitive learning for enhanced audiovisual emotion classification," IEEE Trans. Affective Computing, vol. 3, no. 2, 2012, pp. 184–198.

[3] A. N. Pargellis, H. K. J. Kuo, and C. H. Lee, "An automatic dialogue generation platform for personalized dialogue applications," Speech Communication, vol. 42, 2004, pp. 329–351.

[4] K. Jokinen and K. Kanto, "User expertise modelling and adaptivity in a speech-based e-mail system," in Proc. COLING, 2004, pp. 87–94.

[5] S. E. Brennan and M. Williams, "The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers," J. Memory and Language, vol. 34, no. 3, 1995, pp. 383–398.

[6] M. Swerts and E. Krahmer, "Audiovisual prosody and feeling of knowing," J. Memory and Language, vol. 53, no. 1, 2005, pp. 81–94.

[7] J. C. Lin, C. H. Wu, and W. L. Wei, "Error weighted semi-coupled hidden markov model for audio-visual emotion recognition," IEEE Trans. Multimedia, vol. 14, no. 1, 2012, pp. 142–156.

[8] Y. Wang and A. N. Venetsanopoulos, "Kernel cross-modal factor analysis for information fusion with application to bimodal emotion recognition," IEEE Trans. Multimedia, vol. 14, no. 3, 2012, pp. 597–607.

[9] Y. Chiba, M. Ito, and A. Ito, "Estimation of userfs state during a dialog turn with sequential multi-modal features," in HCI International 2013-Postersf Extended Abstracts, 2013, pp. pp. 572–576.

[10] O. Collignon *et al.* "Audio-visual integration of emotion expression," Brain research, vol. 1242, 2008, pp. 126–135.

[11] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," Int. J. Computer Vision, vol. 91, 2011, pp. 200–215.

 ISBN: 978-1-61208-325-4

Modeling the Determinants of Medical Information Systems Usability in Saudi Arabia

Majed A. Alshamari and Mohamed E. Seliaman
Department of Information Systems
King Faisal University
Hofuf , Saudi Arabia
e-mail: smajed@kfu.edu.sa, mseliaman@gmail.com

***Abstract*—Saudi Arabia's healthcare sector is rapidly moving towards fully automating medical records in all hospitals throughout the country to create the ability to have the medical information move from hospital to hospital as announced in ambitious e-health program. In spite of the wide adoption of IT systems in healthcare sector, very little limited research has been conducted to investigate health and medical information systems perceived usability within the Saudi context. This paper attempts to fill the gap in the literature of medical information systems usability by modeling the usability determinants of Medical Information Systems within the context of Saudi Arabia.**

Keywords-Usability testing; usability measures; medical system

I. INTRODUCTION

The role of Information and Communication Technology (ICT) has shown significant impacts on most aspects of life in Saudi Arabia during the last decades. Both government and private sectors made huge investment in ICT infrastructures and have shown growing trend to adopt ICT as strategic enabler to leverage the efficiency and effectiveness of their processes. Saudi Arabia's healthcare sector as an important example is rapidly moving towards fully automating medical records in all hospitals throughout the country to create the ability to have the medical information move from hospital to hospital as announced in ambitious e-health program under which about 220 hospitals and 2,000 Primary Healthcare Centers (PHCs) will be automated [1][2]. However, there are still several barriers to successful implementation of medical information systems [3][4].

In spite of the wide adoption of IT systems in healthcare sector in many countries, criteria addressing usability are notably absent [5]. In this regards, the health-care industry faces various challenges and great pressures in order to adopt IT. The successful adoption and utilization can greatly lead to reduce process inefficiencies and health-care cost. In addition, health care quality can be also improved. However, having adopted IT in healthcare industry may lead to error-prone and misuse by users "clinicians" [5]. Thus, testing and evaluating healthcare information systems usability can play a vital role in making such systems and reducing errors.

Usability engineering methods tend to measure systems usability. It aims to shorten lifecycle of systems developments, improve quality of the systems and reduce the cost [6][7][8]. The requirements of a certain system usually depend on its characteristics such as medical systems usually aims to achieve users' trust while working under pressure, accommodate users errors and very reliable, safe and accurate [9]. Failing to achieve the expected requirements can cause severe usability problems. For example, having a default value caused serious issues. This has been reported as a failure to enter a new dosage levels as the system did not prompt the user for the data [9]. Another reported example is that the poor usability is the given reason for having critical errors and may result to lethal implications [6][10]. In fact, significant hazard can be caused by poor usability [9]. Hence, it can be seen clearly the importance of having a usable medical system taking into account usability guidelines. However, in order to measure usability properly, usability measures of medical systems should be clearly identified and properly recorded. An extensive review on existing usability measure and models is presented in [11].

The current literature suggests that a "reasonable" usability is acceptable in medical systems, although others systems aim to achieve high level of usability. This can be seen due to the nature of medical systems. These types of systems prioritize different attribute such as safety, accurately and efficient as critical factors, whereas other types of systems such as e-commerce classify user satisfaction, effectiveness and learnability as critical factors [6][10]. The latest published medical usability standards attempt to control safe use. It has been classified as a critical standard for medical system usability [6][9]. Moreover, efficiency and error free use have been described to be success factor of Electronic Health Records (EHRs). Although, there are some commercial medical systems, they can not tailor all the needed specification of a certain clinic or a hospital [6].

From measurement prospective, measuring medical system is a different from measuring other types of system. The reasons are: Firstly, medical system has a different nature of clinical work domain, such as multiple users sometimes are required to perform a task. Secondly, privacy and legal issues can be a significant obstacle [6]. Usability testing is usually recommended to be conducted naturally. In

 ISBN: 978-1-61208-325-4

addition, poor measurement of system functionality and usability may lead to patient injuries and deaths [9]. Even most skilled users can be misled by user's interfaces if they do not follow specific design guidelines [9]. Furthermore, medical devices and systems have to be used effectively and safely, therefore their interaction and design should be considered when design and evaluate [6]. However, recent researches aim to provide usable health systems to enable their users to concentrate on their patients rather than the systems issues [6]. In addition, it has been reported that each dollar spent on usability can offer up to 30 dollar in systems investment. Usability is now a fundamental criterion to buy software [12].

In this paper, we attempt to fill the gap of the literature of medical information systems usability by proposing a set of customized usability measures for medical information systems in Saudi Arabia. Specifically, the objectives of this research are: to explore the current literature of medical information systems in Saudi Arabia, to propose a set of customized measures on medical information systems in Saudi Arabia, and to empirically examine the current usability issues of medical systems in Saudi Arabia through applying customized measures.

II. LITERATURE REVIEW

Recently, there was a growing literature focus on healthcare information systems usability. Viitanen et al. [13] used a national web questionnaire with nearly 4000 physicians actively working in patient care in Finland. They described three dimensions of clinical ICT system usability: compatibility between clinical systems and physicians' tasks, the support for information sharing and collaboration in clinical work. Their results indicated several usability problems and deficiencies which considerably hindered the efficiency of clinical ICT use and physician's routine work.

Kjeldskov et al. [14] conducted a usability evaluation with novice users when an electronic patient record system was being deployed in a large hospital. They repeated the evaluation After 15 months of system usage by the nurses in their daily work. Results show extensive use and experience with systems will not solve usability problems.

Khajouei et al. [15] examined and compared the effectiveness of Cognitive Walkthrough (CW) and Think Aloud (TA) usability evaluation methods, for identifying usability problems. Their study involved two usability evaluators and 10 physicians were recruited to perform usability testing of a CPOE system (Medicator). Results from this study show that there is no significant difference between the performance of the CW and the TA methods in terms of number of usability problems identified and the mean severity of these problems. They recommended a combination of methods is advised as the most appropriate approach for usability evaluation to avoid problems which can lead to potentially fatal consequences.

Karahoca et al. [16] examined usability of two alternative prototypes for medical information systems using Nielsen's heuristic evaluation and cognitive walkthrough methods. Their study is based on a case study of 32 potential users of medical information system prototypes. Their case study results confirmed the view that the usability evaluation results of iconic Graphical User Interfaces (GUIs) were better than those of non-iconic GUIs in terms of Nielsen's heuristic evaluation, effectiveness and user satisfaction.

Jaspers [17] presented an overview of the methodological and empirical research available on the three usability inspection and testing methods most often used for testing interactive health technologies: the heuristic evaluation, the cognitive walkthrough, and the think aloud.

Horsky et al. [18] conducted a research study to characterized and compared four usability evaluation methods used during the design and pilot testing of new clinical documentation software. Their results reported that no single evaluation method outperforms others methods in detecting all or most usability problems.

III. RESEARCH METHOD

This research is intended to identify the critical factors or determinants for measuring the perceived usability of Medical/Health Information Systems. By Medical/Health Information Systems we mean all types of computerized information systems developed for recording, processing, retrieving and managing patients' medical/health information. Based on an extensive review of relevant literature, eight usability measures were selected and included in the research investigation. These measures are: Learnability, Safety, Trustfulness, Usefulness, Effectiveness, Efficiency, Satisfaction, and Productivity.

A. *Measurement Development*

This study is based on the survey questionnaire method for the purpose of collecting the required data. The survey questionnaire tool used consists of three parts: the first part is designed to collect the participants' demographic data, the second part is designed to get the participants' rating of proposed factors, and last part is used to record participants' attitude toward using Medical Information Systems.

The survey questionnaire measurement tool included 34 items forming 8 latent variables. Cronbach's alpha was used for testing the internal consistency reliability of the scale. All constructs reported alpha value above the acceptable threshold of 0.7, except the learnability which reported a value of 0.59 (see table 1) . This construct was removed and excluded from any further analysis. In addition, convergent validity of the constructs was tested using a principal component analysis. Only factors with eigenvalues greater than 1.0 and component loadings exceeding 0.5 were considered significant and hence kept for further analysis.

TABLE 1. RELIABILITY STATISTICS FOR THE CONSTRUCTS

Construct	N of Items	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items
Learnability	4	0.597	0.569
Safety	5	0.798	0.801
Trustfulness	4	0.811	0.811
Usefulness	4	0.726	0.728
Effectiveness	7	0.840	0.842
Efficiency	3	0.822	0.823
Satisfaction	2	0.670	0.670
Productivity	3	0.748	0.748

B. The Study Sample

A total number of (200) forms were distributed to medical staff working in public and private hospitals in Saudi Arabia. Later, (104) survey questionnaire form were returned making a response rate of (52%). Five forms were excluded from any further analysis because of missing data. Hence, the remaining (99) valid filled survey forms were used in the analysis. The females represented a majority of the respondents (75.8%) indicating low participation from males. The majority of respondents aged between 20 and 40 years (77%). The data indicated that 26.3% of the respondents have no experience with any Medical Information Systems at all. In addition, 23 % reported less than one year experience with Medical Information Systems. While 54% of the respondents reported more than 6 years of experience with using computers, only 16% of the sample reported more than 6 years of experience with using Medical Information Systems (see Table 2).

TABLE 2. DEMOGRAPHIC CHARACTERISTICS OF THE RESPONDENTS

Measure	Item	Frequency	Percentage
Gender	Female	75	75.8%
	Male	24	24.2 %
Age	Between 20-30	49	49.5 %
	Between 31-40	27	27.3 %
	Between 41-50	12	12.1 %
	Over 50	9	9.1 %
	Missing	2	2.0 %
Experience with Computer	Less than a year	8	8.1 %
	1 to 6 years	33	33.3 %
	Greater than 6 years	55	55.6 %
	Missing	3	3.0 %
Experience with any HIS	No experience	26	26.3 %
	Less than a year	23	23.2 %
	1 to 6 years	25	25.3 %
	Greater than 6 years	16	16.2 %
	Missing	9	9.1 %

IV. RESULTS AND ANALYSIS

The analysis of this paper is based on average values of the responses for each factor to identify which factors can be considered critical for each usability measure. The factors with means exceeding or equal to 4.11 were recognized as critical factors.

Based on this analysis, factors that were identified as critical for health information systems safety are: (1) personal information should be protected, (2) the system should explain the errors clearly, (3) the system should explain what to do when a problem faces users, (4) resources should be handled without any hazard, (5) the system should maintain a specific level of performance in case of being faulty. According to results from Analysis of Variance (ANOVA), there is no significant difference between female ratings for this measure. But, this analysis reports a significant effect of the flow of experience with using computers on the respondents' rating for the fifths item in this measure (p value=0.02).

The factors that were identified critical for health information trustfulness are: (1) visual and text content should be easily understood, (2) the system should give clear user assistance in its operation, (3) the system should be clear in terms of its purpose and objectives, (4) the user should feel in control of the system product. According to results from analysis of variance (ANOVA), there is no significant difference between female ratings for this measure. Also, there is no significant effect of the flow of experience with using computers or HIS on the respondents' rating for the items in this measure.

The factors that were identified critical for health information usefulness are: (1) the system should provide up to date and complete information, (2) the system should consume appropriate amount and types of resources when it functions, (3) unnecessary elements should be eliminated from the user interface without significant information loss, and (4) the user should customize system interface to his preferences.

The factors that were identified critical for health information effectiveness are: (1) the system should be flexible for users to achieve their work goals and tasks, (2) the system should enable users to complete their work tasks accurately, (3) the system should be consistent in achieving different work tasks, (4) the system should provide the users with feedback on completing their tasks, (5) the system should provide the user with help to solve problems and recover from errors, and (6) the system should enable to complete their tasks with minimal number of errors.

The factors that were identified critical for health information efficiency are: (1) the system should enable users to complete their work tasks timely, (2) the system should enable users to complete their work tasks with the available resources, and (3) the system should enable users to complete their tasks with minimal action.

The factors that were identified critical for health information satisfaction are: (1) the system should be attractive to use and (2) using the system in performing the work tasks is pleasing. Results from analysis of variance (ANOVA) did not report any significant difference between female ratings for this measure. Also, there is no significant effect of the flow of experience with using computers or

Medical Information Systems on the respondents' rating for all items in this measure.

The factors that were identified critical for health information productivity are: (1) the system should increase users' productivity; (2) the system should enable users to complete their work tasks with the available resources, and (3) the system should enable users to complete their tasks with minimal loading time.

In addition to identifying the critical factors of Medical Information Systems usability, the research also investigated the attitudes of the study subjects towards using Medical Information Systems and their behavioral intention to use such systems. The results reported an average rating of (4.2) for the three attitude used in the study. The sample subjects also reported an average of (4.0) as rating for the three variables used to measure their behavioral intention to use Medical Information Systems.

Analysis of Variance (ANOVA) results report a significant effect of flow of experience using computers on the sample reported attitudes towards using Medical Information Systems (p value=0.02).

V. Discussion and Conclusion

In this paper, an extensive literature review is conducted. A set of customized usability measures is defined. A survey questionnaire tool is designed. A pilot study is conducted to assess the content validity, clarity and relevance of the survey questionnaire elements. A revised version is administered to a sample of medical staff working in public and private hospitals. Preliminary results show that all proposed usability measures except those of learnability are considered critical and may influence the user's intention to interact with the health information systems.

In addition, the study findings indicate that the adoption of Medical Information Systems in Saudi Arabia is at its infancy stage; since more than 25% of the respondents have reported that they have no experience with any Medical Information Systems at all and more than 22 % reported they have less than one year experience with Medical Information Systems.

In general, there is no significant effect of the personal characteristics of the study sample on their rating of the importance of the different identified determinants of Medical Information Systems usability measures, with the exception of flow of experience using computers which has shown a significant effect on the item: "the system should maintain a specific level of performance in case of being faulty".

Acknowledgment

The author would like to acknowledge the support for this research provided by Deanship of Scientific Research: King Faisal University as a part of the research grant (#130090).

References

[1] A. Alkraiji, T. Jackson, and I. Murray, "Barriers to the Widespread Adoption of Health Data Standards: An Exploratory Qualitative Study in Tertiary Healthcare Organizations in Saudi Arabia," Journal of medical systems, vol. 37(2), 2013, pp. 1-13.

[2] M. M. Altuwaijri, "Supporting the Saudi e-health initiative: the Master of Health Informatics programme at KSAU-HS," Eastern Mediterranean Health Journal, vol. 16(1), 2010, pp. 119-124.

[3] B. Aldosari, "Rates, levels, and determinants of electronic health record system adoption: A study of hospitals in Riyadh, Saudi Arabia," International Journal of Medical Informatics, 2014, In Press.

[4] M. Khalifa, "Barriers to Health Information Systems and Electronic Medical Records Implementation. A Field Study of Saudi Arabian Hospitals", Procedia Computer Science, vol. 21, 2013, pp. 335-342.

[5] P. Edwards, J. Moloney, J. Jackoc, and F. Sainfort, "Evaluating usability of a commercial electronic health record: A case study," Int. J. Human-Computer Studies, vol. 66, 2008, pp. 718–728.

[6] S. Kujala, "User involvement: a review of the benefits and challenges," Behaviour & information technology, vol. 22(1), 2003, pp. 1-16.

[7] R.G. Bias, D.J Mayhew, Cost-Justifying Usability. Morgan Kaufmann, San Francisco, CA, 1994.

[8] C.G. Wolf, J. Karat, "Capturing what is needed in multi-user system design: observations from the design of three healthcare systems," Proc. the 2nd conference on Designing interactive systems: processes, practices, methods, and techniques (DIS '97), ACM, August 1997, pp. 405-41, Amsterdam, The Netherlands.

[9] S. Braun, "Usability for medical devices," Proc. IEEE Symposium on Product Safety Engineering, Oct. 2005, pp. 16-22, 10.1109/PSES.2005.1529517.

[10] R. Koppel, J.P. Metlay, A. Cohen, B. Abaluck, A. R. Localio, S. E. Kimmel, and B. L Strom, "Role of computerized physician order entry systems in facilitating medication errors," The Journal of the American Medical Association, vol. 293 (10), 2005, pp.1197–1203.

[11] A. Seffah, M. Donyaee, R. B. Kline, and H. K. Padda, "Usability measurement and metrics: A consolidated model," Software Quality Journal, vol. 14(2), 2006, pp. 159-178.

[12] N. Juristo, AM. Moreno, and M.I. Sanchez-Segura, "Analysing the impact of usability on software design," Journal of Systems and Software, vol. 80, issue 9, September 2007, pp. 1506-1516.

[13] J. Viitanen, H. Hyppönen, T. Lääveri, J. Vänskä, J. Reponen, and I. Winblad, "National questionnaire study on clinical ICT systems proofs: Physicians suffer from poor usability," International Journal of Medical Informatics, vol. 80(10), 2011, pp. 708-725.

[14] J. Kjeldskov, M. B. Skov, and J. Stage, "A longitudinal study of usability in health care: Does time heal?," International Journal of Medical Informatics, vol. 79(6), June 2010, pp. 135-143.

[15] R. Khajouei, A. Hasman, and M.W.M. Jaspers, "Determination of the effectiveness of two methods for usability evaluation using a CPOE medication ordering system," International Journal of Medical Informatics, vol. 80(5), May, 2011, pp. 341-350.

[16] A. Karahoca, E. Bayraktar, E. Tatoglu, and D. Karahoca, "Information system design for a hospital emergency department: A usability analysis of software prototypes," Journal of Biomedical Informatics, vol. 43(2), April 2010, pp. 224-232.

[17] M. W.M. Jaspers, "A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence," International Journal of Medical Informatics, vol. 78(5), May, 2009, pp. 340-353.

[18] J. Horsky, K. McColgan, J. Pang, A. J. Melnikas, J. A. Linder, J. L. Schnipper, and Middleton, "Complementary methods of system usability evaluation: Surveys and observations during software design and development cycles," Journal of Biomedical Informatics, vol. 43(5), October 2010, pp. 782-790.

An Embodied Group Entrainment Characters System Based on the Model of Lecturer's Eyeball Movement in Voice Communication

Yoshihiro Sejima (Yamaguchi University)
Graduate School of Science and Engineering,
Computer Science and Design Engineering,
sejima@yamaguchi-u.ac.jp

Tomio Watanabe (Okayama Prefectural University)
Faculty of Computer Science and Systems Engineering,
Department of Systems Engineering,
watanabe@cse.oka-pu.ac.jp

Mitsuru Jindai (University of Toyama)
Graduate School of Science and Engineering,
Mechanical and Intellectual Systems Engineering,
jindai@eng.u-toyama.ac.jp

Atsushi Osa (Yamaguchi University)
Graduate School of Science and Engineering,
Computer Science and Design Engineering,
osaa@yamaguchi-u.ac.jp

Yukari Zushi (Yamaguchi University)
Graduate School of Science and Engineering,
Perceptual Sciences and Design Engineering,
s019ff@yamaguchi-u.ac.jp

***Abstract*—We have developed a speech-driven embodied group entrained communication system called "SAKURA" for enabling group interaction and communication. In this system, speech-driven computer-generated (CG) characters called InterActors with functions of both speakers and listeners are entrained as a teacher and some students in a virtual classroom by generating communicative actions and movements. In this study, for enhancing group interaction and communication, we analyze the eyeball movements of a lecturer communicating in a virtual group by using an embodied communication system with a line-of-sight measurement device. On the basis of the analysis results, we propose an eyeball movement model that consists of a saccade model and a model of the lecturer's gaze at the audience, called "group gaze model." The saccade model reveals eyeball movement with a delay of 0.20 s with respect to the lecturer's head movement. A group gaze model reveals the rate of the lecturer's gaze (Center: 60%, Left-side: 27%, Right-side: 13%). Then, we develop an advanced communication system in which the proposed model is used with SAKURA. Using this system, we perform experiments and carry out sensory evaluation for determining the effects of the proposed model. The results reveal that the proposed model is effective for group interaction and communication in the speech-driven embodied group entrainment characters system.**

Keywords-Human Interface; Human Interaction; Embodied Communication; Group Interaction; Eyeball Movement.

I. INTRODUCTION

With the advancements in the field of information technology, it is now becoming possible for humans to use CG characters called avatars to communicate in a 3D virtual space over a network [1]. Because the avatars express nonverbal behavior based on keyboard commands, current systems do not simulate embodied sharing using the synchrony of embodied rhythms such as nodding and body movements in human face-to-face communication. In such communications, not only verbal messages but also nonverbal behavior such as nodding, body movements, gaze, and facial expressions are rhythmically related and mutually synchronized between talkers [2]. This synchrony of embodied rhythms in communication is called entrainment, and it generates the sharing of embodiment in human interactions [3].

Focusing on the entrainment of embodied communication in our previous work, we analyzed the entrainment between a speaker's speech and a listener's nodding motion in a face-to-face communication and developed InterRobot Technology (iRT), which generates a variety of communicative actions and movements such as nodding and body movements using a speech input [4]. In addition, we developed an interactive CG character called "InterActor" and demonstrated that it can effectively support human interactions and communication. We also developed a speech-driven embodied group entrained communication system called "SAKURA" for enabling group interaction and communication in which InterActors are entrained as a teacher and some students in a virtual classroom. Furthermore, we demonstrated that the developed system could effectively support human interactions and communication [5].

In group communication, not only the lecturer's body movements but also the line-of-sight of the lecturer, such as gaze and eye contact, play an important role in enhancing the embodied interaction and communication [6]. Furthermore, it has been reported that the line-of-sight is important for enhancing the embodied interaction in an avatar-mediated

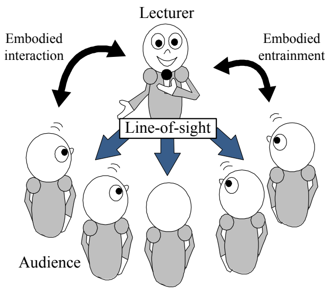


Figure 1. Research concept.

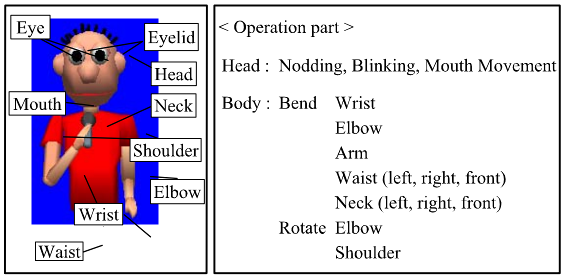


Figure 2. InterActor.

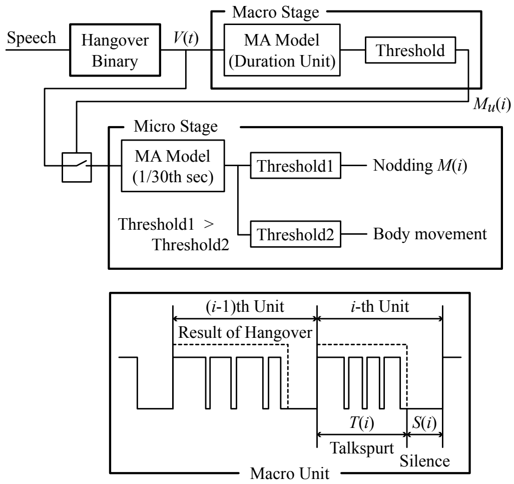


Figure 3. Listener's interaction model.

communication. For example, the teleconference system was developed by using some CG characters with the technology of mixed reality, and demonstrated that the line-of-sight of talkers was influenced by the position and direction of CG characters [7]. In addition, the interactive communication system which controls the line-of-sight among three avatars was developed based on the talker's voice information such as sound pressure and pitch, and demonstrated that the interaction of voice was enhanced by modulating the line-of-sight of the talker [8]. However, it is difficult to enhance such embodied interaction because the characteristics of a lecturer's line-of-sight in group communication have not been established thus far. Therefore, it is essential to develop a group embodied communication system that has the characteristics of a lecturer's line-of-sight in order to enable smooth communication during an embodied interaction [9] (Figure 1).

In this study, we analyze a lecturer's behavior in a virtual group communication. In particular, by focusing on the lecturer's line-of-sight, the eyeball movements are measured by using a line-of-sight measurement device, and the characteristics of the lecturer's line-of-sight such as group gaze are analyzed. On the basis of the analysis results, we propose an eyeball movement model that consists of a saccade model and a model of the lecturer's gaze at an audience, called "group gaze model," for enhancing group interaction and communication. In order to evaluate the effects of the proposed model on group interaction and communication, we develop an advanced communication system in which the model is used with SAKURA. The effectiveness of the proposed eyeball movement model is demonstrated for performing the communication experiments with a sensory evaluation using the developed system.

II. Analysis of Line-of-Sight in Group Communication

A. InterActor

In order to support human interaction and communication, we developed a speech-driven embodied entrainment character called InterActor, which performs the functions of both speaker and listener (Figure 2). The listener's interaction model includes a nodding reaction model that estimates the nodding timing from a speech ON-OFF pattern and a body reaction model linked to the nodding reaction model [4]. The timing of nodding is predicted using a hierarchy model consisting of two stages: macro and micro (Figure 3). The macro stage estimates whether a nodding response exists or not in a duration unit, which consists of a talkspurt episode $T(i)$ and the following silence episode $S(i)$ with a hangover value of 4/30 s. The estimator $M_u(i)$ is a moving-average (MA) model, expressed as the weighted sum of unit speech activity $R(i)$ in (1) and (2). When $M_u(i)$ exceeds a threshold value, nodding $M(i)$ also becomes an MA model, estimated as the weighted sum of the binary speech signal $V(i)$ in (3).

$$M_u(i) = \sum_{j=1}^{J} a(j)R(i-j) + u(i) \tag{1}$$

$$R(i) = \frac{T(i)}{T(i)+S(i)} \tag{2}$$

$a(j)$: linear prediction coefficient

$T(i)$: talksqurt duration in the i - th duration unit

$S(i)$: silence duration in the i - th duration unit

$u(i)$: noise

$$M(i) = \sum_{j=1}^{K} b(j)V(i-j) + w(i) \tag{3}$$

$b(j)$: linear prediction coefficient

$V(i)$: voice

$w(i)$: noise

The body movements are related to the speech input because the neck and one of the wrists, elbows, or arms, or the waist are operated when the body threshold is exceeded. The threshold is set lower than that of the nodding prediction of the MA model, which is expressed as the weighted sum of the binary speech signal to nodding. In other words, when InterActor functions as a listener for generating body movements, the relationship between nodding and other movements is dependent on the threshold values of the nodding estimation.

The body movements in the case of a speaker are also related to the speech input by operating both the neck and one of the other body actions at the timing over the threshold, which is estimated by the speaker's interaction model as its own MA model of the burst-pause of speech to the entire body motion [4]. Because speech and arm movements are related at a relatively high threshold value, one of the arm actions in the preset multiple patterns is selected for operation when the power of speech is over the threshold.

B. *Experimental System*

The experimental setup is shown in Figure 4. In this experiment, for using InterActor as a virtual listener, three isomorphic displays (I･O DATA LCD-AD203G) were used. InterActor was represented with each display having a resolution of 1600 x 1200 pixels; it was generated using Microsoft DirectX 9.0 SDK and a Windows XP workstation (CPU: Corei7 2.93 GHz, Memory: 3 GB, Graphics: NVIDIA Geforce GTS250). The frame rate at which InterActor was represented was 30 fps. The three displays were synchronized using the image distributor (ELECOM VSP-A2). The lecturer and the left and right displays made up one corner each of an equilateral triangle having a side length of 2 m. The positions and angles of the lecturer's body movements were measured using four magnetic sensors (Polhemus FASTRAK) placed on the top of the lecturer's head, both wrists, and the back of the lecturer's body. The image of the lecturer's eyeball was measured using a line-of-sight measurement device [10] (Figure 5) and was input to the PC through an A/D converter (CANOPUS ADVC110). The angle of the lecturer's eyeball movement was calculated by the template matching of the cornea (Figure 6). The voice was sampled using 16 bits at 11 kHz with a headset (SONY DR-260DP). The measured data were recorded on an HDD in real time.

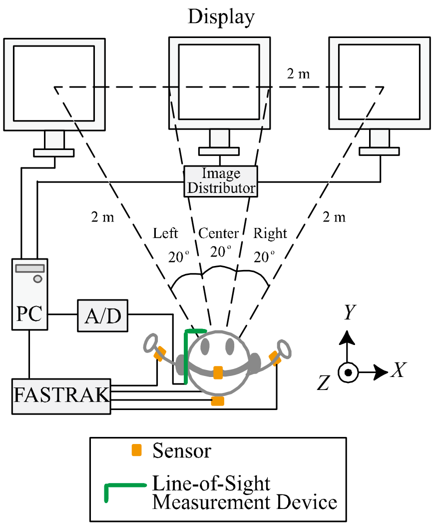


Figure 4. Experimental setup.

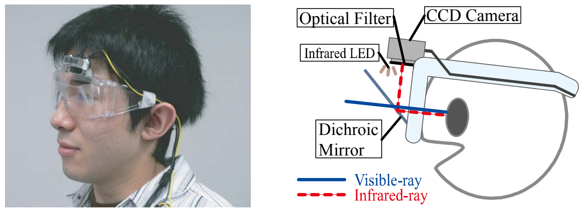


Figure 5. Line-of-sight measurement device.

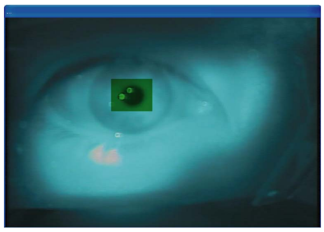


Figure 6. Eyeball image.

Figure 7. Experimental setup.

The experimental process was as follows: First, the lecturer used the system for 1 min for the calibration of his eyeball movement. Next, the lecturer was told to talk on general conversational topics to the three InterActors for 5 min (Figure 7). The conversational topics were not specified, and the instructor was told that the rate of his gaze was to be equally divided between the three InterActors. The three InterActors behaved as virtual listeners by nodding and making body movements in real time. Ten male students played the role of the lecturer.

C. *Analysis of Line-of-Sight in Group Communication*

First, the rate of the lecturer's gaze was analyzed for the three InterActors. In this analysis, using the data on head movement, we defined the lecturer's gaze as Right-side, Center, or Left-side, as shown in Figure 4. An example of the time change of the head movement is shown in Figure 8. The figure shows that the lecturer mainly gazed at the center and the duration for which the lecturer subconsciously gazed at the Left-side was longer than that for the Right-side. The average duration of the lecturer's gaze is given in Table 1. The average duration of the Center gaze accounted for approximately 60% of the time, and the duration of Left-side (27%) was twice that of the Right-side gaze (13%). This

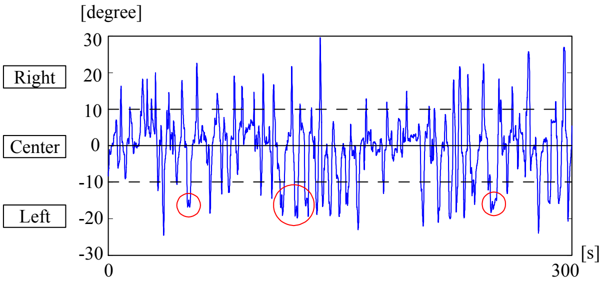


Figure 8. Example of time change of head movement.

result is consistent with the fact that the Left-side is

Table 1. Result of gaze duration.

	Left	Center	Right (%)
Average	27.6	59.5	12.9
Standard deviation	20.0	24.9	8.7

predominant over the Right-side in the spatial cognitive functions of humans [11] [12].

Next, the relationship of the timing between the lecturer's head movement and the eyeball movement during the communication experiment was analyzed. The angle of Z axis $p(i)$ was measured at a sampling rate of 30 Hz using the magnetic sensor placed on the lecturer's head; this angle was then used for calculating the head movement $x(i)$ by measuring the difference between the following and the previous data $[p(i+1)-p(i-1)]$. The coordinate value $c(i)$ was measured at the sampling rate 30 Hz using the line-of-sight device on the lecturer's head; $c(i)$ was then used for calculating the eyeball movement $y(i)$ by taking the difference between the following and the previous data $[c(i+1)-c(i-1)]$. The relationship of time changes was analyzed using the following cross-correlation function $C(\tau)$.

$$C(\tau)=\frac{\sum_{i=1}^{n-\tau}\{x(i)-\mu_x\}\{y(i+\tau)-\mu_y\}}{\sqrt{\sum_{i=1}^{n}\{x(i)-\mu_x\}^2}\sqrt{\sum_{i=1}^{n}\{y(i)-\mu_y\}^2}} \qquad (4)$$

μ_x, μ_y : average of x, y

n : number of data

τ : time delay

Figure 9 shows an example of time change of the cross-correlation function $C(\tau)$ in an analysis period of 30 s over 5 min. A strong positive correlation between the lecturer's head movement and eyeball movement was confirmed for $\tau = 0.20$ s. The average delay time of the eyeball movement was 0.196 ± 0.06 s over 5 min. This result showed that the eyeball movement had a delay time with respect to the head movement of 0.20 s during group communication; this delay time was similar to the latent time of a saccade in the body functions of a human (0.20 s) [13] [14].

The obtained results can be summarized as follows: The lecturer mainly gazed at the center, and he gazed at the Left-side twice as much as the Right-side. Further, the eyeball movement had a delay time with respect to the head movement of 0.20 s.

III. EYEBALL MOVEMENT MODEL

In this research, we have proposed an eyeball movement model that generates a lecturer's eyeball movement for enhancing group communication on the basis of the characteristics revealed by the abovementioned analysis. This model consists of a saccade model and a model of the lecturer's gaze at an audience called the "group gaze model." An outline of the proposed model is as follows:

A. *Saccade model*

The main characteristic of the saccade model is an eyeball movement with a delay of 0.20 s with respect to the lecturer's head movement. First, the angle of the lecturer's head movement was calculated in a virtual space. If the lecturer's head moved, the eyeball moved with a delay of 0.20 s with respect to the head movement in the same direction (Figure 10).

B. *Group gaze model*

The characteristic of the group gaze model is the duration of the lecturer's gaze (Center: 60%, Left-side: 27%, Right-side: 13%) [9].The lecturer's gaze is generated stochastically with an exponential distribution based on the abovementioned analysis. An example of the lecturer's gaze

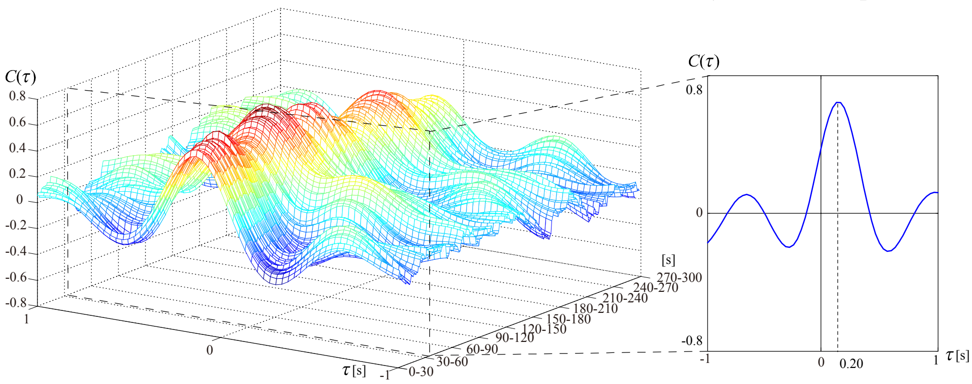


Figure 9. Example of time change of cross-correlation $C(\tau)$ between a lecturer's head movement and his own eyeball movement.

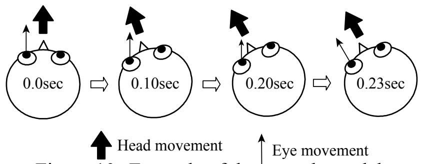


Figure 10. Example of the saccade model.

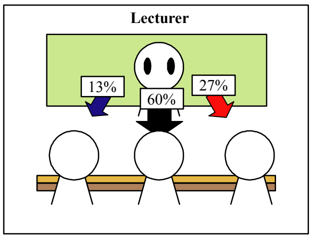


Figure 11. Group gaze model.

is shown in Figure 11. It is expected that the lecturer's gaze will be effective for group communication when the gaze duration is varied.

IV. A SPEECH-DRIVEN EMBODIED GROUP ENTRAINMENT CHARACTERS SYSTEM

A. SAKURA

A speech-driven embodied group entrained communication system called "SAKURA" has been developed for enabling group interaction and communication [5]. Five InterActors play the role of students, and one InterActor plays the role of a teacher; they are arranged in a virtual classroom where they are entrained on the basis of only a speech input. By using SAKURA, talkers can communicate with a sense of unity through the entrained InterActors by using only a speech input.

B. Speech-driven embodied group entrainment system

An advanced communication system was developed in which the proposed model was used with SAKURA (Figure 12). The virtual space was generated by Microsoft DirectX 9.0 SDK and a Windows XP workstation (HP workstation xw4200: CPU: Pentium4 2.8 GHz, Memory: 1 GB, Graphics: NVIDIA Quadro FX3400). The voice was sampled using 16 bits at 11 kHz. The frame rate at which the CG characters were represented was 30 fps.

When a lecturer's speech is fed into the system as an input, the lecturer's InterActor generates communicative body movements and actions, and an eyeball movement based on the proposed model. An example of the communicative actions and movements of the lecturer is

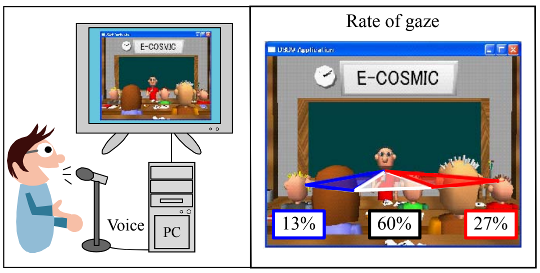


Figure 12. Configuration of the developed system.

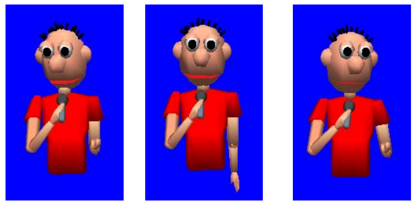


Figure 13. Example of InterActor's body motion and gaze as lecturer.

shown in Figure 13. The mouth operation of the InterActor was realized by the switching operation synchronized with the burst-pause of voice. The audience characters respond to the utterances with appropriate timings by means of their entire body movements, including nodding, blinking, and communicative actions in the manner of listeners [4]. As a result, the lecturer's InterActor gives the audience a natural line-of-sight and generates a communication environment in which the sense of unity is shared by the embodied entrainment.

V. EVALUATION OF EYEBALL MOVEMENT MODEL

A. Evaluation experiment of saccade model

A preliminary experiment was performed to evaluate the saccade model. The main characteristic of a saccade model is an eyeball movement having a delay of 0.20 s with respect to the lecturer's head movement. We performed an experiment in which the delay time of the eyeball movement had three patterns: 0.00 s, 0.20 s, and 0.40 s. An experimental communication scene using the system is shown in Figure 14. Ten male subjects were used for evaluating the three patterns. The result showed that the most common delay was 0.20 s; it was observed in eight out of 10 subjects. Therefore, in this research, the delay time for the eyeball movement was set to be 0.20 s.

B. Evaluation experiment of group gaze model

The effectiveness of the saccade model was demonstrated in the foregoing section. In this section, the group gaze model is evaluated.

1) Experimental Method: In this experiment, the subjects were asked to watch a video. The video was made by recording the display using a system input to the recorded speech data in 2 min. The speech content was an opinion on consumption tax. In this experiment, three modes were compared: in the first mode, a lecturer gazed at the center of the audience (mode (A)); in the second, the lecturer gazed at the entire audience equally (mode (B))


Figure 14. Subject watching video to evaluate modes.

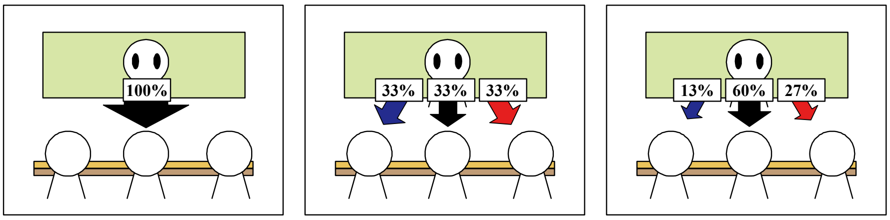


Figure 15. Outline of each mode in the communication experiment.

[15]; and in the third mode, the lecturer gazed at the audience by using the group gaze model (mode(C)) (Figure 15).

The experimental procedure was as follows: first, the subjects watched the video in each mode. Next, they were instructed to perform a paired comparison of the modes. In the paired comparison experiment, based on their preferences, they selected the better mode. Finally, they watched the video in each mode and evaluated each mode on a seven-point bipolar rating scale ranging from -3 (not at all) to 3 (extremely); 0 denoted moderation. The subjects were 30 Japanese students (15 females and 15 males) presented with the abovementioned three modes in a random order.

2) Results of Sensory Evaluation: The result of the paired comparison is shown in Table 2. Further, Figure 16 shows the calculated results of the evaluation provided in Table 2 and based on the Bradley-Terry model given in (5) and (6) [16].

$$p_{ij} = \frac{\pi_i}{\pi_i + \pi_j} \quad (5)$$

$$\sum_i \pi_i = const.(=100) \quad (6)$$

π_i : Intensity of i

p_{ij} : probability of judgment that i is better than j

The consistency of the matching of the modes was confirmed by performing a test of goodness of fit ($x^2(1,0.05) = 3.84 > x_0^2 = 0.02$) and the likelihood ratio test ($x^2(1,0.05) = 3.84 > x_0^2 = 0.02$). The proposed mode (C) was evaluated to be the most affirmative, followed by the (B) equal-gaze and (A) center-gaze modes.

Table 2. Result of paired comparison.

	(A)	(B)	(C)	Total
(A)		5	2	7
(B)	25		7	32
(C)	28	23		51

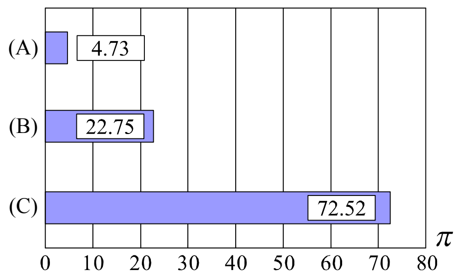


Figure 16. Comparison of π.

The questionnaire result is shown in Figure 17. From the results of the Friedman signed-rank test, we found that "Interaction," "Natural line-of-sight," "Unification," "Realistic sensation," "Vividness," and "Preference," had a significance level of 1% between mode (C) and mode (A). Further, "Lecturer's gaze," had a significance level of 1%, and "Interaction," and "Preference," were at 5% between mode (C) and mode (B). In both the experiments, mode (C)

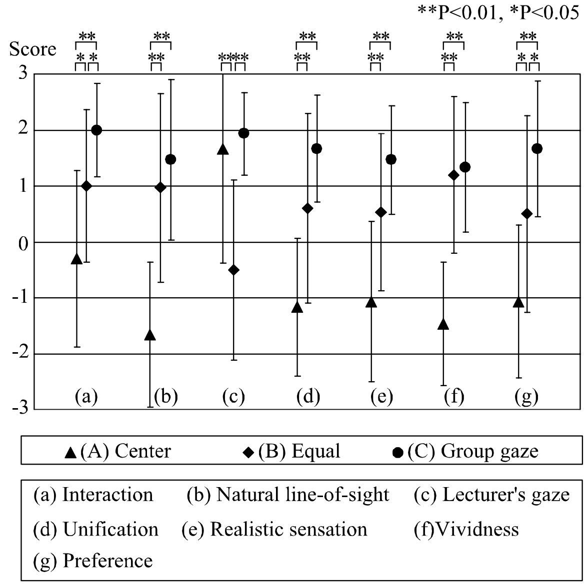


Figure 17. Seven-point bipolar rating.

of the proposed eyeball movement model was most often evaluated to be the best with respect to group interaction and communication. These results indicated that the proposed model in which a lecturer gazed at the center of the audience moderately (60%) was the best of the models considered in this study.

C. Evaluation of group gaze rate

The effectiveness of the group gaze model in which a lecturer gazed at the center of the audience 60% of the time was demonstrated in the foregoing section. In this section, the gaze rate of the group gaze model is evaluated by

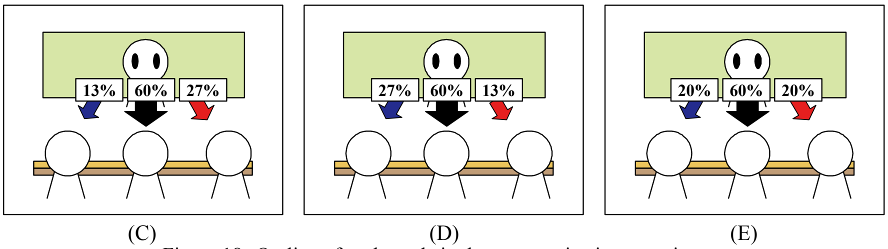


Figure 18. Outline of each mode in the communication experiment.

changing the gaze rate of the remaining 40% of the gaze duration.

1) Experimental Method: In this experiment, three modes were compared: in the first, a lecturer gazed at the audience with the group gaze model (mode (C)); in the second, a lecturer gazed at the audience for reversing the gaze rate between the right-side and the left-side (mode (D)); and in the third mode, a lecturer gazed at the audience for equalizing the gaze rate between the right-side and the left-side (mode (E)) (Figure 18). The experimental procedure was the same as that detailed in Section V.B.1. The subjects were 30 Japanese students (15 females and 15 males) other than the ones mentioned in the foregoing section.

2) Results of Sensory Evaluation: The result of the paired comparison is shown in Table 3. Further, Figure 19 shows the calculated results of the evaluation provided in Table 3.

The consistency of the matching of the modes was confirmed by performing a test of goodness of fit ($x^2(1, 0.05) = 3.84 > x_0^2 = 1.40$) and the likelihood ratio test ($x^2\ (1, 0.05) = 3.84 > x_0^2 = 1.41$). The proposed mode (C) was evaluated to be the most affirmative, followed by the (E) equal-gaze and (D) reverse modes.

The questionnaire result is shown in Figure 20. From the results of the Friedman signed-rank test, we found that "Unification," had significance level of 5% between mode (C) and mode (D). Further, "Unification," had a significance level of 5% between mode (D) and mode (E). In both the experiments, mode (D) of reversing the gaze rate between

Table 3. Result of paired comparison.

	(C)	(D)	(E)	Total
(C)		24	14	38
(D)	6		10	16
(E)	16	20		36

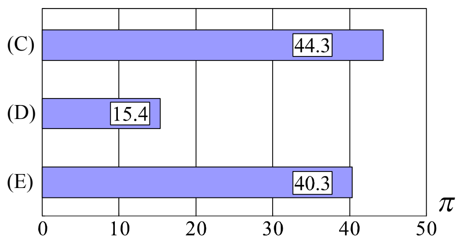


Figure 19. Comparison of π.

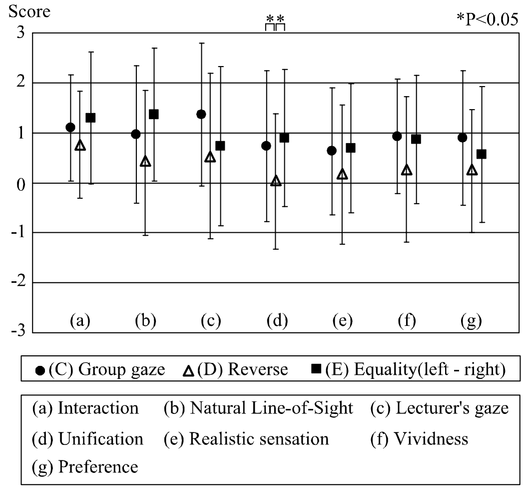


Figure 20. Seven-point bipolar rating.

the right-side and the left-side was not evaluated for group interaction and communication. This result is consistent with the theory that the left-side is predominant over the right-side in the case of the spatial cognitive functions of humans [11] [12].

D. Effectiveness of group gaze model

In the foregoing section, mode (D) in which the gaze rate was reversed between the left-side and the right-side was not evaluated with respect to group communication. In this section, the effect of group gaze is evaluated by comparing mode (C) to mode (E).

1) Experimental Method: In this experiment, two modes were compared: in the first, a lecturer gazed at the audience using the group gaze model (mode (C)); in the second mode, a lecturer gazed at the audience for equalizing the gaze duration between the right-side and the left-side (mode (E)). The experimental procedure was the same as that detailed in Section V.B.1. The subjects were another 20 Japanese students (10 females and 10 males).

2) Results of Sensory Evaluation: The result of the paired comparison is shown in Figure 21. The questionnaire

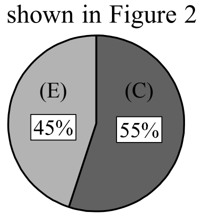


Figure 21. Result of paired comparison.

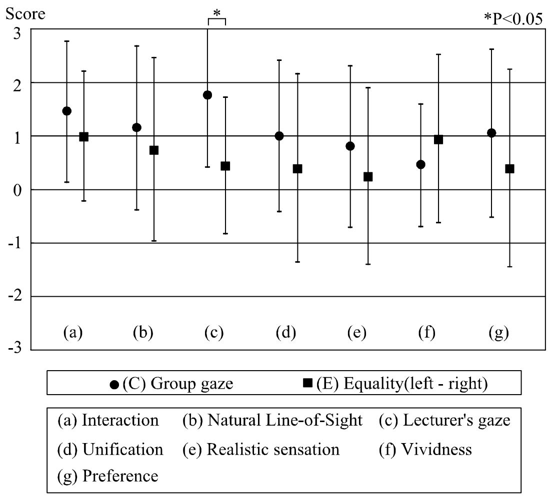


Figure 22. Seven-point bipolar rating.

result is shown in Figure 22. In accordance with the results of the Wilcoxon signed-rank test, the parameter "Lecturer's gaze," had a significance level of 5% between mode (C) and mode (E). This result indicates that the system with the proposed model is effective in group interaction and communication.

VI. CONCLUSION

In this study, we analyzed the characteristics of a lecturer's eyeball movement in the case of group communication by using an embodied communication system with a line-of-sight measurement device. On the basis of the analysis results, we proposed an eyeball movement model that consists of a saccade model and a model of the lecturer's gaze at an audience, called "group gaze model," for enhancing group interaction and communication. The proposed model could be summarized as follows: the main characteristic of a saccade model is an eyeball movement with a delay of 0.20 s with respect to the lecturer's head movement, and a group gaze model has the following durations of the lecturer's gaze: Center 60%, Left-side 27%, and Right-side 13%. Further, we developed an advanced communication system in which the proposed model was applied to a speech-driven embodied group entrained communication system. By using this system, we performed communication experiments and carried out a sensory evaluation. The effectiveness of the proposed model was demonstrated during group interaction and communication in a speech-driven embodied group entrainment characters system.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 24700536, 22300045, 24118707, 25330239.

REFERENCES

[1] Dilley, C., "Virtual World Summit 2007 Report," ASCII, 2008, pp. 94-97.

[2] Condon, W.S. and Sander, L.W., "Neonate movement is synchronized with adult speech,"Science, no. 183, 1974, pp. 99-101.

[3] Watanabe, T., "Human-Entrained Embodied Interaction and Communication Technology for Advanced Media Society,", Proceedings of 16th IEEE International Symposium on Robot and Human Interactive Communication(RO-MAN2007), 2007, pp. 31-36.

[4] Watanabe, T., Okubo, M., Nakashige, M., and Danbara, R., "Interactor: Speech-driven embodied interactive actor,"International Journal of Human-Computer Interaction, vol. 17, 2004, pp. 43-60.

[5] Watanabe, T. and Okubo, M., "SUKURA: Voice-Driven Embodied Group-Entrained Communication System," Proceedings of the 10th International Conference on Human-Computer (HCI International 2003), vol. 2, 2003, pp. 558-562.

[6] Kendon, A., "Some Functions of Gaze-direction in Social Interaction," Aca Psychologica, vol.26, 1967, pp. 22-63.

[7] Inoue, T., "Mixed Reality Meeting System Enabling User to Keep and Share Interpersonal Distance in the Real World," Transactions of Information Processing Society of Japan, vol.50, no.1, 2009, pp. 246-253.

[8] Ishii, R., Miyajima, T. and Fujita, K., "Avatar's Gaze Control to Facilitate Conversation in Virtual-Space Multi-User Voice Chat System, " Transactions of Human Interface Society, vol.10, no. 1, 2008, pp. 87-94.

[9] Sejima, Y., Watanabe, T., Jindai, M., Osa, A. and Zushi, Y., "A Speech-driven Embodied Group Entrainment System with the Model of Lecturer's Eyeball Movement," Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN2012), 2012, pp. 1086-1091.

[10] Sejima, Y., Watanabe, T. and Jindai, M., "An Embodied Communication System Using Speech-Driven Embodied Entrainment Characters with an Eyeball Movement Model," Transactions of the Japan Society of Mechanical Engineers, Series C, vol.76, no.762, 2010, pp.340-350,.

[11] Kato, T.,"Psychology of a Sense of Space (in Japanese),"Shin-Yo-Sha Ltd., 2001, pp.37-47.

[12] Nehira, K., "The Left and Right of Animate Nature (in Japanese)," KYORITSU SHUPPAN CO. LTD., 1998, pp.103-106.

[13] Fukushima, J., "Disturbance of Eye Movements in Schizophrenic Patients," Bulletin of Hokkaido University Collection of Scholarly and Academic Papers, vol.2, 1989, pp.1-13.

[14] Ohno, T., "What Can Be Learned From Eye Movement?: Understanding Higher Cognitive Processes From Eye Movement Analysis," Transactions of the Japanese Cognitive Science Society, vol.9, no.4, 2002, pp.565-579.

[15] Aramaki, M., "Skill of Presentation (in Japanese)," SANNO Institute of Management Publication dept., 2010, pp.39-42.

[16] Luce, R. D., "Individual Choice Behaviour: A Theoretical Analysis," New York: J. Wiley, 1959.

Persona Usage in Software Development: Advantages and Obstacles

Jane Billestrup, Jan Stage
Research Centre for Socio-Interactive Design
Department of Computer Science
Aalborg University, Denmark
{jane, jans}@cs.aau.dk

Lene Nielsen, Kira S. Hansen
Center for Personas Research & -Application
Games & Interaction Design
IT University of Copenhagen, Denmark
{lene,kist}@itu.dk

Abstract—***The Personas technique has been promoted as a strong tool for providing software developers with a better understanding of the prospective users of their software. This paper reports from a questionnaire survey regarding knowledge about Personas and their usage in software development companies. The questionnaire survey was conducted in a limited geographical area to establish the extent of Personas usage within all companies in the chosen region and determine whether software development companies used Personas during the development process. Several issues were identified as reasons for either not using the technique or for poor application of it. The study showed that 55% of the respondents had never heard about Personas. Among those who had heard about the Personas technique, the following obstacles towards usage of the technique were identified: Lack of knowledge of the technique, lack of resources (time and funding), Sparse descriptions – when applied and Personas not being integrated in the development.***

Keywords—Personas, software development, questionnaire survey, grounded theory

I. INTRODUCTION

The common understanding of the Personas technique is, that a Persona is a description of a fictitious person [6, 21] based on data. The main way to represent a Persona is as a text describing the fictional user and a photo depicting the fictional user.

Looking at the related work four issues stand out: 1) it is reported that software developers lack knowledge and understanding of their users, their work, and goals [1], [3] 2) the Personas technique has been promoted as a strong tool for providing the software developers with a better understanding of the potential users [7] 3) several papers conclude that the use of Personas has been a success [10], [12] 4) the Personas technique is not necessarily an incorporated part of the toolbox in the software development industry [15] and the industry might have problems using Personas [2]. Our study will be related to these four issues in the discussion.

The literature provides several examples about what will happen if the Persona technique are used to its full potential. However, there are still unanswered questions about what it means if developers are not using Personas to represent their user groups. Could this lead to: 1) software developers being unfamiliar with the user groups' actual needs 2) software without relevant features 3) less profit than if the software was developed to its full potential. Unfortunately, the literature does not provide much information about these speculations. Other aspects of Persona usage are also unexplored as so are whether other techniques are used instead of Personas.

The purpose of this paper is to explore to what extent software development companies use Personas and whether the industry uses the technique as proposed in the literature. The paper reports from a study in a defined region in Denmark and is part of a larger study on how the Personas technique is applied by software developing companies. This paper reports if companies in this defined region have had success using Personas and incorporating the technique as a part of their development toolbox. The greatest advantage of using one defined region is that it is possible to establish contact with all companies located in the region, which gives a more complete picture than picking out companies located in several regions or countries.

The following section presents a more detailed description of the work related to this study. It describes how Personas are constructed and used, including the pitfalls to avoid. Section 3 presents the methods used for data collection, which consisted of an online questionnaire with both open and closed questions. Grounded theory and coding were used for analysing the qualitative content of the questionnaire. Section 4 presents the results from the questionnaire. Section 5 provides a discussion of the results in a broader context. Finally, section 6 provides the conclusion.

II. RELATED WORK

The literature offers four different perspectives regarding Personas [18]: 1) Cooper's goal-directed perspective 2) Grudin, Pruitt and Adlin's role-based perspective 3) The engaging perspective, which emphasises how the story can engage the reader. These three perspectives agree that the Persona descriptions should be founded on data. However, 4) the fiction-based perspective, does not include data as a basis for Persona description, but creates Personas from the designers' intuition and assumptions. Even though the

Personas technique has been around for more than a decade, when comparing the four perspectives, it is still unclear what and how much background material is required to create Personas [17]. The common perceived benefits of Personas are two-fold: 1) when designing products the technique facilitates that designers remember that they differ from the end-users and 2) the technique enables designers to envision the end-user's needs and wants. Furthermore, in the design process the Personas increase the focus on users and their needs. The technique is an effective communication tool, which uses the Persona description to acquire direct design influence and lead to better design decisions and definition of the product's feature set [6], [7], [12], [13], [14], [16], [21].

Problems have been reported regarding creation and distribution of the developed Personas [2], [22]. The descriptions have been perceived as unreliable and not well communicated. Also developers lacked understanding of how to use the Personas [2], [21], [22]. The technique itself is criticised for being too founded on qualitative data and as a consequence of that - being non-scientific, being difficult to implement, not being able to describe actual people as it only portrays characteristics, and for preventing designers from meeting actual users [1]. Moreover, the unsolved question about how many users one Persona can represent is perceived as problematic [5].

Some have tried to prevent poor use of the Personas technique, e.g. Faily and Flechais [11] describes regularly sending information about the Personas to the team, to ensure that the designers and developers consider the Personas in the design process. They s suggested that the creators should hand over instructions and provide tools that support the developers' usage [11]. Problems in application are reported as also incorporating the mindset of the developers, which is documented by both Blomquist and Arvola [2] and Pruitt and Adlin [21].

In line with this, Matthews et al. [15] focussed mainly on designers and user experience professionals who had some training in Personas creation and had done extensive work with Personas used them as described by others [6], [21]. These designers had a very positive attitude towards the technique. Those who had done minor use of Personas had a moderate or neutral opinion regarding Personas, and those who had not worked with Personas at all had a negative or indifferent opinion regarding the technique.

III. METHOD

In order to study the usage of Personas we conducted a questionnaire study in 60 companies in a limited geographical area. We chose to focus on a well-defined geographical area in order to allow us to do as complete a survey with as many companies as possible. We made considerable efforts to identify and contact all companies in the area. Had we chosen a larger area, the selection of companies would be more random. The aim of the study was to identify the level of knowledge about the Personas technique in a large proportion of the software companies in that area.

A. Participants

We focused on companies that were developing software, either for internal or external use. We ended up with software companies with the following characteristics:

The company develops software with a graphical user interface (e.g., mobile phones, games, web applications, PC or PDA software). The company develops software for customers or for internal use and is geographically located within a limited geographical area. The company employs more than a single person and it is not a hobby company.

To obtain a list with as many software development companies as possible we acquired two lists containing software companies located in the chosen region. These lists were from a previous study of companies and an industry network. This was followed by a search on Linked-In to include companies that only had a smaller development department in the region and had their headquarters located either in another region or in another country. Table 1 shows the process of obtaining the total amount of 134 software companies in the region, which was within the scope of this study.

TABLE I: OBTAINING A LIST OF RELEVANT COMPANIES

	Companies		
Lists to find the companies	**Total number of companies on list**	**Out of scope or gone out of business**	**Applicable companies**
List 1	*77*	***-35***	***42***
List 2	*139*	***-63***	***76***
Linked-in	*16*	***0***	***16***
Total number of applicable companies			***134***

B. Data Collection

We created an online questionnaire using the tool SurveyXact [26]. The first part of the questionnaire was made to gain more information about the respondent and his or her place of employment (e.g. job function, business, number of employees in the company and line of business, within software development). The second part was designed to acquire knowledge of if respondents knew what a Persona was and what it was used for. The third part was about the use of Personas in the companies. This part was only filled out by the people who answered that they knew of, and worked with, Personas. The questionnaire consisted of 35 questions, but only respondents who knew of- and who is working with Personas in their current employment got to answer all 35 questions. The questionnaire consisted of both open and closed questions.

The distribution of the questionnaire was done in two ways. First, 43 companies in which we had a known contact person was contacted by phone. Then the remaining 91 of the 134 companies were contacted to acquire a contact person. This ended in 112 emails being sent out with a link to the questionnaire. Of the 22 companies that we did not send an email, eight had declined to participate and for the rest we could not locate a viable phone number or email address. The recipients were given three weeks to fill out the questionnaire survey. The data collection process resulted in 69 responses in total or a response rate of 51.5%. Of the 69 respondents nine did not finish the questionnaire, leaving us with 60 completed responses.

C. Data Analysis

Grounded Theory was chosen as analysis method because of its ability to help building theory, meaning that we could focus on keeping an open mind during the data collection process..

The aim of grounded theory is described as "*building theory, not testing theory*" [20]. This means that theory should emerge while the analysis takes place and should not be used to prove an already existing theory. Data analysis was conducted continuously while the questionnaire was still open for submissions, as described by Urquhart [24]. When the questionnaire was closed, the data was updated with the results from the latest incoming questionnaires. In the questionnaire we used both open and closed questions. To quantify the open questions, the grounded theory approach, as described by Corbin and Strauss [8], Urquhart [24] and Urquhart et al. [25], was used as an analysis method. All responses to closed questions was analysed quantitatively.

1) Open Questions

Coding was used to analyse the open questions. One question was: "*How would you explain what a Persona is and how it is used?*". For this question the following coding categories were assigned: technique (for creating Personas), finding target user group, when in the process the Personas are used and how they are used. Grounded theory coding was not used for other open questions since the respondents mainly answered in very short sentences and they were sent directly to the end of the questionnaire when they answered "No". E.g. "*Have you ever heard about Personas*?" or "*Have you ever worked with Personas?*" meaning that the number of respondents dropped for every question. As it makes no sense to ask a respondent about their knowledge about the use of Personas if they have already indicated they have never heard about Personas.

2) Closed Questions

Statistics was conducted directly from the closed questions. Results from both open and closed questions are being described in the following section.

IV. RESULTS

This section presents the results. The section is divided into two sub-sections. *Knowledge about the Personas technique* is referring to the first part of the questionnaire. This subsection reports if the Personas technique has been adopted by the software developing companies in the defined region. The second subsection *The understanding of Personas and their use* is dividing the obstacles of Personas usage into four main areas.

A. Knowledge about the Personas technique

The results of the questionnaire indicate that 27 out of 60 respondents, or 45%, have heard about Personas. Fifteen respondents out of 60 have worked with Personas. Seven respondents out of 60 are using Personas as a development tool in their current job. Meaning that 11.5% of the responding companies are currently using Personas as a development tool and 55% of the respondents have never heard about the technique. The distribution across different sizes of companies is shown in Table 2. In this table the dispersion across company size and the number of respondents familiar with Personas.

TABLE II: DISTRIBUTION ACROSS COMPANIES' SIZE

	Number of employees				
Number of companies	1-10	11-50	51-200	>200	**Total**
Using Personas	1	3	1	2	**7**
Not using Personas	23	16	8	6	**53**
Total	**24**	**19**	**9**	**8**	**60**

In Table 3 the 53 responding companies that do not use Personas have been grouped. It shows that 33 respondents have never heard about Personas. Three of the organisations did use Personas at some point but stopped.

TABLE III: RESPONDENTS' KNOWLEDGE ABOUT Personas – FROM COMPANIES THAT DOES NOT USE THEM.

	Number of employees				
Knowledge about Personas	1-10	11-50	51-200	>200	**Total**
Never heard about Personas	18	7	6	2	**33**
Have used Personas, but stopped	2	1	0	0	**3**
Heard about Personas, but don't use them	4	5	2	2	**13**
Worked with Personas in other employment or while studying	1	2	0	1	**4**
Total	**25**	**15**	**8**	**5**	**53**

One respondent stated they used Personas in a project where they collaborated with a group of university students, but did not find the Personas technique useful for other projects. The other two respondents stated that their respective companies stopped using Personas because they did not find the developed Personas applicable in their line of development. 13 respondents stated they had heard about the Personas technique but had never worked with creating Personas themselves and four respondents had worked with creating Personas in an earlier employment or while studying.

B. Understanding of Personas and their use

An open question in the questionnaire was applied coding to reveal all the participating companies' understanding of the term “Persona”. “Personas being an imaginary user”, were expressed by 22 respondents. e.g. “*a fictitious user of the system you are developing*”. “Personas are used as a validation of the design”, were expressed by 17 respondents. e.g. “*making sure user needs are met by a given design*”.

A Persona “being a representation of a larger user segment” was expressed by 13 respondents, e.g. “*description of a set of characteristics characterizing a certain group of users' behavioural patterns*”. And Personas “being a tool for making sure to keep the users and their needs in mind all the way through the development process” were recognised by four respondents. e.g. “*...the Personas are used as focus points for planning the entire product life cycle*”. This means that Personas by far are recognised as fictionalised users used as a tool for designing features requested by users and user segments. On the other hand no more than four respondents expressed that Personas should be used through the entire development cycle. This means that the common idea seems to be that Personas are mainly a tool for identifying some aspects of the user group and not so much a tool to be used during the entire development process.

1) Lack of Knowledge (of the technique)

Lack of knowledge about the Personas technique seems to be a major obstacle regarding usage of Personas as shown in table 3. The analysis showed that 55% of the respondents had never heard about the concept or technique. Of the respondents who had never heard about Personas, 10 people were CEOs, owners or partners (primarily in micro- or small sized companies), five were managers in IT and three worked as sales managers (all three in medium sized companies). In table 4 the respondents job titles has been divided into smaller groups based on whether the company is currently working with Personas, or not.

This indicate that the chance of allocating resources to Personas development might be slim. One respondent indicated that the company did not recognise the importance for any communicative tools. “*The company has downsized and has eliminated the communications position since it is primarily a production company and they don't really understand the importance of e.g. Personas, ambassadors, first movers, e.g. or communication in general for that matter*”. This means that in these companies the knowledge about the Personas technique will not come from management, and even if employees bring the knowledge about Personas into the companies funding will probably not be allocated. On the other hand as seen in table 4, in the seven companies currently working with Personas four respondents was CEO, CTO or owner.

TABLE IV: DISTRIBUTION OF JOB TITLES OF RESPONDENTS

Job function of respondents	Not working with Personas in current employment	Company currently working with Personas
CEO, CTO, Owner	12	**4**
System Developer / Consultant	11	**1**
Project / Product / Sales Manager	16	**0**
Business architect / Communication and PR	8	0
UX / Web Designer / Manager	6	2
Total	**53**	**7**

2) Lack of Resources (time and funding)

The analysis found that Personas are mainly created if a need has been localised for a specific project and “cutting a corner” when using Personas seems to be the general idea. Some only use Personas to the point that they think it creates value for the customer and thereby, profit for the company. Also, when asked in the survey how much resources were allocated to develop Personas, the general answer was zero.

3) Sparse descriptions

When a Persona is created too superficially the Persona will lack the depth that would normally be the strength of the technique, making the Personas untrustworthy and unusable. This contradicts with what helps making Personas useful tools that lead to better design decisions [6], [13], [14], [16], [21]**.** When a Persona is created with much detail and described as a whole character, and not a stereotype, it will support the design and innovation process.

One respondent indicated difficulty in finding a suitable template for the descriptions and that they wanted to create short descriptions instead of detailed character descriptions. “*It is hard to find good templates for constructing Personas. We ended up with a few lines in bullets describing each Persona, which could be used as a fast reference. Instead of a large scheme describing lots of details nobody wanted to read anyway*”. This corresponds with the descriptions of

Personas by some respondents answering the questionnaire. These descriptions were quite superficial and did not describe individual Personas but mainly a job role and a use situation.

4) Not integrated in the development

This ties-in with the finding of lacking resources. The superficial Personas are created to be used in the design process. The descriptions are not meant to be used in any other stages of the design process. Furthermore, they are not used to keep reminding neither developers nor designers about the end-user's and their needs. This means that the potential of the Personas technique is not explored.

C. Advantages of using Personas

The respondents currently using Personas described why their companies are using Personas as follows: "*to support the development of a system that is easy to use for types of user...It is very important for us that the system will be very easy to use, which is why a mapping of the various user groups is important"*. Another respondent stated: "I*nternally in the company Personas are used to communicate characteristics of the customer segments that we want to focus on especially.* Yet another respondent stated that Personas are primarily used for "*optimizing the product"*. These advantages corresponds with the advantages identified in the related work section.

V. Discussion

Next, the results will be discussed in relation to the four issues located in the related work: 1) software developers lack knowledge and understanding of their users, their work, and goals [1], [3] 2) the Personas technique has been promoted as a strong tool for providing the software developers with a better understanding of the potential users [7] 3) several papers conclude that the use of Personas has been a success [10], [12] 4) the Personas technique is not necessarily an incorporated part of the toolbox in the software development industry [15] and the industry might have problems using Personas [2].

A. Lack of knowledge and understanding of the users

Software developers lack knowledge and understanding of their users, e.g. their work and goals [1], [3]. Among our findings was poor application of the technique. This relates perfectly to the first point about developers lacking knowledge and understanding of the users, since the Personas' descriptions, if applied, are made sparse and only used in a very narrow time frame of the development process. One of our findings was that the development of the Personas lacked resources, since none of our respondents had a budget allocated specifically for the Personas development. This goes against the related work stating that Personas can lead to better design decisions [6], [12], [13], [14], [16], [21].

B. Personas can help developers understand users

The Personas technique has been promoted as a strong tool for providing software developers with a better understanding of the potential users [7]. In our questionnaire and the section in this paper describing the advantages of using Personas, it was indicated that the most useful thing when using the Personas technique was that Personas helped the team share a specific and consistent understanding of several, different user groups. Which can lead to another advantage of product optimization.

C. Personas used as a successful tool

Several papers conclude the use of Personas has been a success [10], [12]. This corresponds with the experiences of our respondents who are using Personas. The tool is described as useful to help developers understand the users and their needs, especially if the system needs to be usable for several different types of end users. Some respondents that are using Personas identified mos challenges for creating Personas. e.g. "*it can be hard to find templates for creating Personas."* another respondent stated that "*it is a challenge to map all user groups without asking all costumers"*. These obstacles has to be solved before Personas can be applied as a useful tool.

D. Personas are not incorporated in the industry

The Personas technique is not necessarily an incorporated part of the toolbox in the software development industry and the industry might have problems using Personas [2]. Since only 44% of our respondents have even heard about the Personas technique and less than 12% have worked with creating Personas, it is fair to say that Personas are not an integrated tool in the software development industry in this region. Also, we found that only four respondents indicated that Personas should be used through the entire development process, meaning that even if Personas are used, they are not necessarily used to their full potential. In companies using Personas, the technique is used mainly to identify types of users or use cases.

The Personas are kept to a minimum and not focused on describing whole characters. As in the related work we found developers lacking understanding of how to use Personas to gain most from their usage [2], [7], [22]. The reasons for that could be a combination of several aspects. We found that resources are not allocated specifically for creating Personas, which corresponds with the area of usability in general [1], [22], [23].

The full potential of Persona usage does not seem to have caught on in the industry. Matthews, Judge and Whittaker [15] found a connection between on the one hand the perception of Personas and on the other hand to what extent the technique was used and the amount of training the developers had had using Personas.

 ISBN: 978-1-61208-325-4

VI. Conclusion

The purpose of this paper was to explore to what extent Personas were used by software development companies in a limited geographical area and whether they used Personas as proposed in the literature. To accomplish this, we conducted a questionnaire survey with usable responses from 60 software development companies.

The study showed that only 7 out of the 60 software development companies used Personas. The results from the questionnaire also uncovered four issues. Lack of knowledge of the technique as such and lack of resources both related to companies not using the Personas technique. Sparse or badly designed descriptions or not being part of the development process both related to poor application, when using the technique.

Our findings are well linked to other studies described in the related work section. Yet our study contributes with a new angle by focusing on making a complete study within a limited geographical area we now have a pretty good idea about if the Personas technique is an integrated tool in software development in this geographical area. We have not been able to find related work that has done a similar study in another country. This means that this paper is the first paper indicating whether Personas are used for developing software at the industry.

The main limitation on our results is that we focussed on a limited geographical area. This was necessary to achieve a high level of coverage of all companies in that area. The alternative would have been less coverage by including the same number of companies but from a larger geographical area. As future work it would be interesting to learn more about the advantages of using Personas. This area still needs further study even though some advantages have been identified in this paper, also it would be interesting to learn if companies that do not use Personas are using another tool instead.

Acknowledgment

We would like to thank the companies and employees that participated in our questionnaire survey and Infinit for supporting the research.

References

[1] J. Bak, K. Nguyen, P. Rissgaard and J. Stage, "Obstacles to usability evaluation in practice: a survey of software development organizations", Proc. of NordiCHI, 2008, pp. 23-32.

[2] Å. Blomquist and M. Arvola, "Personas in action: Ethnography in an Interaction Design Team", Proc. of NordiCHI, 2002, pp. 197-200.

[3] A. Bruun and J. Stage, "Training software development practitioners in usability testing: an assessment acceptance and prioritization", Proc. of ozCHI, 2012, pp. 52-60.

[4] Y. Chang, Y. Lim and E. Stolterman, "Personas: From Theory to Practices", Proc. of NordiCHI, 2008, pp. 439-442.

[5] C. N. Chapman and R. Milham, "The Personas' new Clothes: Methodological and Practical Arguments Against a Popular Method, Proc. of HFES, 2006, pp. 634-636,.

[6] A. Cooper, "The inmates are running the asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity", Sams Publishers, 1999.

[7] A. Cooper and R. Reimann, "About face 2.0: The Essentials of Interaction Design", Wiley Publishing, 2003.

[8] J. Corbin and A. Strauss, "Basics of Qualitative Research. Techniques and Procedures for Developing Grounded Theory", 3rd edition, Sage Publications, 2008.

[9] H. W. Desurvire, "Faster, Cheaper!! Are Usability Inspection Methods as Effective as Empirical Testing?", John Wiley & Sons, 1994, pp. 173-202.

[10] A. Dotan, N. Maiden, V. Lichter and L. Germanovich, "Designing with Only Four People in Mind? - A Case Study of Using Personas to Redesign a Work-Integrated Learning Support System", Proc. of INTERACT, 2009 pp. 497-509.

[11] S. Faily and I. Flechais, "Persona Cases: A Technique for Grounding Personas", Proc. of CHI, 2011 pp. 2267-2270.

[12] J. Grudin and J. Pruitt, "Personas, Participatory Design and Product Development: An Infrastructure for Engagement", Proc. of PDC. 2002 pp. 144-161.

[13] F. Long, "Real or Imaginary - the Effect of Using Personas in Product Design." In IES Conference, Dublin: Irish Ergonomics Review, 2009 , pp. 1-10.

[14] J. Ma, and C. LeRouge, "Introducing User Profiles and Personas into Information Systems Development", AMCIS . Vol. 237, 2007.

[15] T. Matthews, T. Judge and S. Whittaker, "How Do Designers and User Experience Professionals Actually Perceive and Use Personas?", Proc. of CHI, 2012 pp. 1219-1228.

[16] T. Miaskiewicza and K. A. Kozarb, "Personas and User-Centered Design: How Can Personas Benefit Product Design Processes?", Design Studies 32, 5. 2011, pp. 417–430.

[17] L. Nielsen, "Personas. In The Encyclopedia of Human-Computer Interaction", 2nd Ed., Aarhus, Denmark: The Interaction Design Foundation, http://www.interaction-design.org/encyclopedia/Personas.html, 2013. (November 15th 2013)

[18] L. Nielsen, "Personas - User Focused Design", Human-Computer Interaction. Springer, 2012.

[19] L. Nielsen and K. S. Nielsen, "Usage of Personas in Danish Companies" – http://Personas.dk/wp-content/Brugen-af-Personas-hos-virksomheder-i-Danmark-20131.pdf, 2013. (November 1st 2013)

[20] S. Pace, "A Grounded Theory of the Flow Experiences of Web Users", Proc. of IJHES, 2003, pp. 327-363.

[21] J. Pruitt and T. Adlin, "The Persona Lifecycle: Keeping People in Mind Throughout Product Design", Morgan Kaufman, 2006.

[22] J. Pruitt and J. Grudin, "Personas: Practice and Theory", Proc. of DUX, 2003, pp. 1-15.

[23] D. Svanæs, and J. Gulliksen, "Understanding the Context of Design – Towards Tactical User Centered Design", NordiCHI, 2008, pp. 353-362.

[24] C. Urquhart, "Grounded Theory for Qualitative Research: A Practical Guide", Thousand Oaks, California: Sage, 2013.

[25] C. Urquhart, H. Lehmann and M. D. Myers, "Putting the Theory Back Into Grounded Theory", guidelines for grouned theory studies in information systems". Info Systems J 20, 2010, pp. 357-381.

[26] www.survey-xact.com

Effects of Wind Source Configuration of Wind Displays on Property of Wind Direction Perception

Width of Wind Velocity Distribution and Accuracy of Wind Source Alignment

Takuya Nakano, Yuya Yoshioka and Yasuyuki Yanagida
Dept. Graduate of Science and Technology
Meijo University
Nagoya, Japan
e-mail: {{123430042, 113430024}@ccalumni., yanagida@}meijo-u.ac.jp

***Abstract*—We examined the property of wind direction perception at the frontal region of the head to find a guideline for optimal wind source arrangement. In previous experiments, localized winds produced by a single compact fan were used as stimuli. Such a localized wind is rather different from the natural uniform wind in a real environment. Because the subjects might be able to judge the wind direction based on the facial region hit by the wind, the performance of discriminating the direction of a localized wind may be different from that for a uniform wind. Thus, in this study, we examined the human ability to discriminate the wind direction using a uniform wind that covered the entire face and compared the result with that for a localized wind. We measured the Just Noticeable Difference (JND) in wind direction perception and found that there was a significant difference between the JND for a uniform wind and that for a localized wind.**

Keywords-Wind sensation; Sensory property; JND

I. INTRODUCTION

Recently, in the technical field of Virtual Reality (VR), systems that reproduce virtual environments using "wind sensation" have been developed. Wind sensation refers to a combination of sensations related to feeling the wind, which is considered to be a kind of haptic sensation [1]. By presenting non-contact stimulation to a user through a wind sensation, we expect the system to provide the user with a greater sensation of presence. Therefore, numerous systems incorporating wind displays have already been developed. For example, there are attractions in amusement parks that present a wind or fragrance with images and sound, e.g., "Soaring" [2] at Disney World. Movie theaters incorporating winds and scents, called "4DX," are now in operation [3]. In addition, to provide the sensation of existing in a target environment, several studies using wind sensations have been conducted. For example, Suzuki *et al*. developed a system that provides air-jet-driven force feedback through a ladle-like handheld tool, which achieved an interaction with force feedback in an untethered manner [4]. Minakuchi proposed the use of wind gusts to help users notice and understand information such as determining the location of an information source based on the wind direction, where the importance of the information is represented by the air volume [5]. Sawada *et al*. developed BYU-BYU-View, which is an input/output interface that uses wind. When a user exhales toward a special screen, a wind emerges from another user's screen. Thus, the user can utilize an application or communicate with a partner [6]. Furthermore, systems such as Windcube [7] and Immersive 3D Wind Display [8] have realized wind presentations from different directions by arranging multiple wind sources around a user. They showed by questionnaires that the simultaneous presentation of a movie and wind using these systems enhanced the sensation of immersion compared with the presentation of a movie alone.

However, in such studies, only a small number of wind sources have been used. Thus, it is unclear whether a precise wind direction can be reproduced using these systems. If the wind sources are arranged too sparsely, it is difficult to precisely reproduce the wind direction in the virtual environment. In contrast, if the wind sources are arranged too densely, users would not be able to discriminate the wind produced by neighboring wind sources, which is regarded as over-engineering. By taking the human ability to discriminate the wind direction into account when designing a wind source configuration, such over-engineering can be prevented, and a natural wind can be reproduced.

We examined the properties of wind direction perception to find a guideline for optimizing the wind source configuration when designing a system that reproduces an environment using wind sensation. In our previous studies[9][10], however, the results might have been affected by errors in the fan alignment (mounting angle) and the variance of the wind velocity distribution generated by each fan, because multiple fixed fans were used in the experimental setup. Furthermore, the localized wind generated by a simple fan was quite different from natural wind. In this study, we found a new guideline for wind source configuration by measuring the property of wind direction perception, using a uniform wind blowing on the entire face to prevent error due to the wind source.

 ISBN: 978-1-61208-325-4

The rest of this paper is structured as follows. Section 2 describes related and previous work. Section 3 presents the results of an experiment and discusses the apparatus and method. Finally, Section 4 concludes this paper.

II. RELATED WORK

Several studies that reproduce an environment using wind sensation have been conducted. A historic example is the famous Sensorama simulator [11] by Heilig. Sensorama was a game in which a player felt like they were riding a motorcycle. Sensorama provided not only visual and auditory stimuli but also tactile and olfactory stimuli, including a wind sensation provided by a wind blowing from the front.

More recently, Moon *et al*. showed that the sensation of presence can be improved by providing wind in addition to showing a movie of a snowstorm [7]. Kosaka *et al*. [8] arranged 25 fans on a dome-shaped frame at intervals of 45°: eight fans on each of three levels (ear-height, 45° higher and 45° lower) and one at the top of the dome. They showed a movie of a person swinging with and without presenting the wind. As a result, the presentation of wind achieved a higher sensation of reality than that without wind. Cardin *et al*. mounted eight fans on a head mounted display at intervals of 45° and presented wind with a movie of a flight simulator [12]. Lehmann *et al*. conducted an experiment to evaluate the sensation of reality for three conditions: presenting only a movie of snow flurries, the movie with wind from a ventilator mounted on a traverse system, and the movie with wind from two fans mounted on a worn helmet. The results showed that 75% of the subjects reported that the ventilator provided the greatest increase in realism [13]. Matsukura *et al*. developed a system that gave a user the sensation that an odor was emanating from a certain position on the screen, by producing winds generated by fans at four corners of the screen and allowing them to collide in front to form a wind in a direction orthogonal to the original direction [14]. They first made vertical winds produced by four fans located at the corners of the screen collide two by two to form two horizontal winds (leftward and rightward) along the screen. They then made these two horizontal winds collide to produce a wind heading toward the user sitting in front of the screen. Hirota *et al*. developed a system that could provide variable wind direction. They presented wind from two fans set in different directions and let them collide obliquely. By controlling the velocity of the wind from each fan, they succeeded in presenting an intermediate wind between two directions [15].

Some researchers have studied the property of wind sensation perception. Kubota *et al*. reported that the wind perception threshold of the face was about 0.2 m/s, depending on the temperature or fluctuation of the flow velocity [16]. Kojima *et al*. investigated which part of the head was sensitive to wind stimulation, in the context of their wearable wind display for local skin stimulation [17]. Their results showed that the regions around the ears were the most sensitive. Hashimoto *et al*. examined the perception of wind at the fingertips [18]. They measured the absolute threshold using the limit method and the difference threshold using the constant method. They also measured the difference threshold of the directional perception. However, they did not conduct tests at the face.

Among these studies, we have focused on the human properties related to perceiving the wind direction. We examined the property of wind direction perception at the frontal region of the head. As a result, values for the Just Noticeable Difference (JND) in wind direction perception were obtained, but a significant inter-subject difference was observed. In this study, we found the possibility that subjects might discriminate the wind direction based on the area of the face touched by the wind [9]. Next, we conducted an experiment to examine the effect of the stimulation area of the face by measuring JNDs using multiple conditions for the stimulation area. We found a significant difference between JND values of wind discrimination for different conditions [10].

In these studies, a single fan was used to provide the wind, following many existing systems with wind presentation. However, natural wind is not a local phenomenon but a uniform sensation encompassing the entire face. Because the point that the wind hits affects wind direction discrimination, we supposed that the human performance of discriminating the wind direction might be different, depending on whether a localized or uniform wind was presented. Another problem in our previous study was that we used multiple fixed fans to present wind from various directions. This might have caused variations in the wind velocity and a slight alignment error for each fan. The slight alignment error might have affected the results because of the potential to discriminate the wind direction by the area hit by the wind. For example, when the wind hit the center of the face, it also hit the left side of the face if we misaligned the position of the fan located on the right side of the subject. Because we arranged the fans one by one by hand, there was a possibility that slight misalignments accrued.

Thus, in this study, we measured the JND values of wind direction perception at the frontal region of the head for both localized and uniform winds. In addition, we compared the JND obtained by using multiple fixed fans to that obtained by moving a fan, to assess the effect of the fan alignment error.

III. EXPERIMENT

A. Apparatus

We used DC fans (SST-AP121 by SilverStone Technology Co. Ltd.; 120 mm^2) as wind sources, following the procedures used in our previous studies [9][10]. The maximum air flow was 1.0 m^3/min, and the operating noise was 22.4 dBA, measured 80 cm from the fan.

We selected this model of fan because it produces an airflow in a circular fashion using a fan filter and swirl-shaped fan grille and is well-suited to examine the properties of wind direction perception because it has better directivity than an ordinary fan. In Fig. 1, we show the airflow difference between this fan and an ordinary fan. The "ordinary" fan used for comparison was a DC fan (109R1212H102 by Sanyo Denki Co. Ltd.; 120 mm^2).

Figure 1. Air flow generated by fan (left: ordinary fan, right: SST-AP121)

The wind velocity distribution generated by SST-AP121 is listed in Table I. We measured wind velocities at intervals of 5 cm within the limits of a 20-cm square on a plane normal to the axis of the fan at a distance of 80 cm from the fan. The left, right, top, and bottom of the table correspond to the directions of wind movement. The unit is m/s. For a comparison with the ordinary fan, we show the wind distribution by 109R1212H102 in Table II. Both fans were driven at the rated voltage (12 V).

TABLE I. WIND DISTRIBUTION BY SELECTED FAN: SST-AP121 (UNIT: m/s)

	Far left	Left	Center	Right	Far right
Uppermost part	0.258	0.350	0.500	0.520	0.422
Upper part	0.576	0.862	1.102	1.062	0.692
Center	0.672	1.146	1.282	1.196	0.680
Lower part	0.416	1.176	1.282	1.066	0.582
Lowermost part	0.326	0.540	0.656	0.546	0.346

TABLE II. WIND DISTRIBUTION BY "ORDINARY" FAN: 109R1212H102 (UNIT: m/s)

	Far left	Left	Center	Right	Far right
Uppermost part	1.516	1.358	1.031	0.678	0.427
Upper part	1.798	1.658	1.308	0.923	0.700
Center	1.705	1.712	1.600	1.238	1.198
Lower part	1.306	1.413	1.381	1.376	1.372
Lowermost part	0.767	0.893	0.869	1.011	1.403

In Table II, the wind spreads to a wider area. In Table I, the wind is the strongest at the center, and the wind velocity becomes smaller in the outer regions. Looking at Fig. 1 and Table II, we can find that the ordinary fan produces wind in an oblique direction rather than straight ahead.

In previous studies [9] [10], we used a single fan as the wind source. In this study, however, we needed to present a uniform wind that could cover the entire face. According to Table II, a single "ordinary" fan was not sufficient to present a uniform wind because the ratio of the minimum to maximum wind velocity was only 0.201.

We designed wind source configurations to present a uniform wind. We found that the "selected" model of fan (SST-AP121) could produce a unimodal wind distribution, i.e., the wind velocity was larger at the center and smaller at the periphery. Therefore, by composing a 2 × 2 or 3 × 3 fan array, we expected to be able to present a uniform wind that could cover the entire face. The composed fan arrays as wind sources are shown in Fig. 2, and the measured wind velocity distributions produced by the fan arrays are listed in Tables III and IV, respectively.

Figure 2. Configurations of fans (left: 2 × 2, right: 3 × 3)

TABLE III. WIND DISTRIBUTION BY 2 × 2 FAN ARRAY (UNIT: m/s)

	Far left	Left	Center	Right	Far right
Uppermost part	0.914	1.358	1.116	0.954	0.920
Upper part	1.080	1.756	1.600	1.482	1.384
Center	1.364	1.862	1.798	1.752	1.452
Lower part	1.442	1.704	1.720	1.798	1.436
Lowermost part	0.920	1.180	1.342	1.792	1.370

TABLE IV. WIND DISTRIBUTION BY 3 × 3 FAN ARRAY (UNIT: m/s)

	Far left	Left	Center	Right	Far right
Uppermost part	1.658	1.796	1.650	1.662	1.598
Upper part	1.876	1.820	1.776	1.870	1.848
Center	1.754	1.834	1.930	1.868	1.788
Lower part	1.904	1.920	1.960	1.940	1.898
Lowermost part	1.704	1.942	1.964	1.892	1.754

From Tables III and IV, the minimum to maximum wind velocity ratios of the 2 × 2 and 3 × 3 fan arrays are 0.491 and 0.840, respectively. In other words, as the number of fans increased, the wind became more uniform. Based on this result, we decided that a 3 × 3 array was sufficient as a wind source unit to present a uniform wind within an area of 20 cm^2, which was a suitable size to cover the head. The unit size was 360 mm × 360 mm. When only the central fan in this unit was activated, it corresponded to the experimental condition of the previous studies [9] [10], in which a single fan was used as a wind source.

In previous studies [9] [10], we placed 13 fans in a range of –60° to +60° with respect to a subject at intervals of 10°. The wind from the front of the face was 0° and was presented from seven fans in a range of –30° to +30° at 1.3 m/s. The separation and limit of this stimulus were based on a prior study where we estimated the wind direction perception. The wind from the front of the subject (0°) was the standard stimulus, and winds from seven positions ranging from –30° to +30° were comparison stimuli. The JND of the wind direction perception was measured using the method of constant stimuli (Fig. 3). However, when using multiple fans, individual fan differences related to the mounting angle and wind distribution might cause a problem.

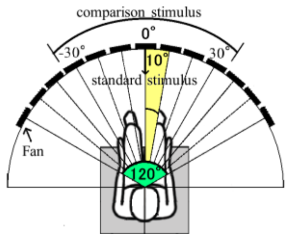


Figure 3. Experimental setup for measuring JND

To prevent individual fan differences from affecting the results, we used the single wind source unit shown in Fig. 2, instead of 13 separate fans. We attached this wind source unit to a moving platform that could be moved on an arc rail whose center was aligned at the center of the subject's head. With this experimental setup, we could eliminate the factor indicated in a previous study [9], where a slight misalignment of the fans could significantly affect the user's perception of the wind direction.

The distance between the wind source and the subject was 80 cm, and the wind velocity was 1.3 m/s. According to Tables I and IV, the newly configured wind source unit could provide wind with a faster velocity than that produced by a single fan, when it was operated at the rated voltage of the fan (12 V). Therefore, to make the wind velocity equal to that produced by a single fan, the wind source unit was operated at 8 V. The unit size was 360 mm × 360 mm, which corresponded to an angle of 25°.

We used a step motor unit (ASC66AK-N5 by Oriental Motor Co. Ltd.) to drive the platform carrying the wind source unit. The maximum rotation velocity was 360 rpm, and the maximum velocity of the platform was approximately 7.6°/s.

To prevent subjects from identifying the direction by the motor noise, we provided white noise using a portable audio player (Walkman NW-754 by SONY Co. Ltd.), along with noise-canceling earphones. Although we needed to prevent subjects from determining the fan location visually, to expose the maximum possible amount of skin area on the face to the wind, we did not use blinders. Subjects closed their eyes throughout the experiment. Fig. 4 shows an overview of the experimental setup.

Figure 4. Overview of experimental setup

To prevent any misalignment of the head position during the experiment, we used a chin support. We separated the small mount for the chin support from the table on which the rail was placed to prevent vibrations caused by the motor and gear from being transmitted to the subjects.

B. Method

We measured the JND using a constant stimulus method. Wind from 0° in front of a subject was the standard stimulus. Wind from any of seven positions within −30° to +30° at intervals of 10° was the comparison stimulus. The subject sat in front of the wind source and put their face on a chin support. The experiment started with their eyes closed.

First, the wind source presented the standard stimulus. Next, it was moved to any one of the positions from −30° to +30° and presented a comparison stimulus. The subject determined whether the comparison stimulus was to the left or right with respect to the standard stimulus. Next, the wind source returned to the first position (0°), and the standard stimulus was presented again, followed by the presentation of a comparison stimulus from any one of the positions from −30° to +30°. We repeated this procedure until each comparison stimulus (−30° to +30°) was presented 10 times, i.e., we presented 140 stimuli in total, including standard stimuli.

Before starting the experiment, we tested whether a subject could hear the motor noise to prevent them from using sound to discriminate wind directions. If the subject could hear the motor noise, they were asked to turn up the volume of the white noise until they could no longer hear the motor.

The duration of each stimulus was 4.5 s, followed by an interval of 8 s, including the moving time for the wind source. The duration for a session was approximately 30 min. A subject was asked to attend two sessions: localized wind and uniform wind, in random order. We allowed a break of approximately 20 min between sessions to let the subject relax and cool the motor. Hence, the total duration of the experiment for one subject was approximately 80 min. Ten male subjects in their twenties volunteered for the experiment.

C. Results

Examples of plots showing the ratios of the times that subjects answered "right" for each comparison stimulus angle are shown in Figs. 5 and 6. Fig. 5 is an example of the results for a single fan, and Fig. 6 shows the results for the 3 × 3 fan array. We fitted a cumulative normal distribution curve and calculated JND values. The JNDs calculated for 10 subjects and a pilot study are listed in Table V.

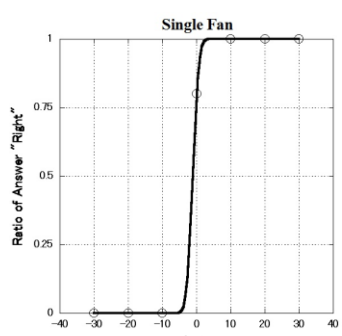


Figure 5. Ratio of times subject answered "right" (single fan)

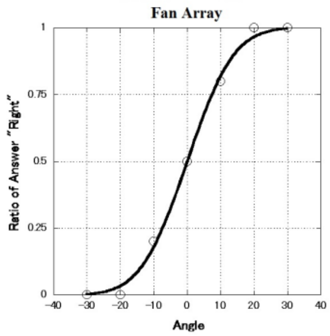


Figure 6. Ratio of times subject answered "right" (fan array)

TABLE V. JNDS OF ALL SUBJECTS (°)

Subject	Pilot study	Single	Unit
1	7.44	4.75	9.17
2	7.44	0.97	5.47
3	3.37	0.94	7.38
4	6.37	1.00	6.12
5	4.97	1.02	1.01
6	6.56	1.00	3.18
7	5.28	4.19	5.28
8	3.37	1.02	5.16
9	8.21	0.87	5.71
10	7.44	1.02	7.04

The average JND value for 10 subjects was 6.05° in the previous study, with values of 1.68° for a single fan and 5.55° for a fan array. The JND was larger for uniform wind. Eight subjects gave all correct answers except for 0° when using a single fan, whereas only one subject could give all correct answers when using a fan array. The standard deviations over the subjects were 1.73° for the previous study, 1.48° for a single fan, and 2.25° for a fan array.

Fig. 7 shows the average JND values, including the previous study [10]. "Pilot study (single)" refers to the results of the previous experiment with a single fan (but using different fans for different directions). "Single" shows the results when moving a single fan, and "3 × 3" shows the results with the fan array. The vertical axis is JND (in degrees). We conducted a t-test[19] to verify whether there was a significant difference between pairs of JND values. First, we conducted a t-test between the results of the previous study [10] and this study with a single fan.

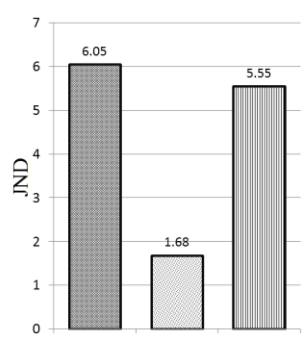


Figure 7. JND for each condition

Because the subjects in the past study and this study were not paired (they used two different groups of subjects), we conducted an unpaired t-test. The two-sided p-value was 0.00001, and there was a significant difference ($p < 0.001$). Because the subject groups of "single fan" and "fan array" were identical, we conducted a paired t-test. The two-sided p-value was 0.00026. Thus, there was a significant difference ($p < 0.001$).

D. Discussion

The results of this study using a single fan were different from those of the previous study [10]. The cause of this difference was considered to be either the alignment error of each fan or individual performance differences in the fans used to present winds from different directions in the previous study. Moreover, in the previous study, the fans were manually fixed on an arc-shaped frame, which might have affected the results. In this study, because we presented wind by moving a wind source unit on semicircular rail, there was little error from the mounting angle and individual fan differences. Thus, we obtained stable results. In the comparison of the results by a single fan and fan array, we found that it was more difficult for subjects to discriminate wind directions when a uniform wind was presented.

Although the JND of subject five for a uniform wind was smaller than that for the local wind, this consisted merely of the difference in the ratio of answering "right" for 0°. Except for 0°, the ratio of correct answers by this subject was 100% for both the local and uniform winds. For the other nine subjects, the JND values for the local wind were smaller. With the localized wind, subjects could discriminate wind directions based on the areas reached by the wind. With the uniform wind, it was difficult for subjects to discriminate the wind direction by the areas reached by the wind, because it blew on their entire face. Therefore, some subjects reported that they changed their strategy to discriminate the wind direction, i.e., focusing on the direction of airflow on the face.

The standard deviation for the fan array was larger than that for the single fan. This implies that subjects could discriminate wind directions using the areas touched by the winds for the local wind, whereas for the uniform wind, the methods used to discriminate wind directions differed among individuals. For the local wind, all of the subjects reported that they discriminated directions using the areas touched by the winds. For the uniform wind, three subjects discriminated directions by the side of the face where the wind flowed. Because the winds produced by the fan array had a larger range than those of the single fan, some subjects reported that they used their neck or arms as a reference for discrimination. In this way, individual differences are subject to occur when a uniform wind is used.

Moreover, differences in the level of concentration existed as individual differences common to the local and uniform winds. The subjects became sleepy in 15 out of 20 sessions (10 subjects, 2 conditions), because their eyes were closed throughout the session. A sensitive subject could detect the motion of the wind source unit by a slight variation in the illumination on his face. Two subjects noticed that the motor vibration was uncommonly transmitted. However, they noted that they could not discriminate the direction using this cue.

IV. Conclusion

In this study, we measured the JND of the wind direction perception of a local wind produced using a single fan commonly used for a wind display and a uniform wind produced by a fan array that was similar to natural wind. As a result, we found that the uniform wind provided larger JND values and made it more difficult to discriminate the wind direction compared to the local wind. By comparing the results with those of the previous study [10], we found that errors in the mounting angle and individual fan performance differences significantly affected wind direction perception. Thus, when designing a wind display, the following guidelines were obtained: (1) if one wants to present a uniform wind, the interval of the wind sources can be larger than when using a single fan, and (2) the fans should be arranged with extremely high precision if multiple fans are used to present winds in different directions.

The average ratios for the subjects answering “right” for 0° was 55% with a single fan and 41% with the wind source unit. It was thought that the center of the wind source was accurately aligned with the subject’s face. However, because of the need for a strict fan configuration, the entire apparatus should also be arranged with extremely high precision.

In this study, the JND for a uniform wind was 5.55°, whereas the size of the wind source unit corresponded to approximately 25°. This means the size of the wind source unit was larger than the JND requirement. In this study, to present a natural wind, we used a 3 × 3 fan array as a wind source unit. Therefore, if we designed a wind display based on a fixed fan array, we would have to arrange fans with a sufficient density to cover all directions. However, we still believe that it is inappropriate to apply the JND value obtained in this study immediately, because there are numerous factors that affect the wind direction perception, e.g., the wind velocity, wind temperature, gender, and age. In this study, the subjects only concentrated on discriminating the wind direction. However, the existence of a movie or sound in an actual wind display might reduce the level of concentration on the wind. Thus, we might be able to make the wind sources sparser than indicated by the results of this study. In future work, we will further investigate the required precision when considering the property of wind perception to achieve acceptable wind displays for people regardless of age and gender. In addition, we are going to examine cross-modal effects with the other sensory modalities.

References

[1] T. Ogi and M. Hirose, “Effect of Multisensory Integration for Scientific Data Sensualization,” JSME International Journal. Ser. C, Dynamics, Control, Robotics, Design and Manufacturing, June 1996, pp. 411–417.

[2] M. W. Summer, “Amusement Apparatus and Method” US Patent 6,354,954, March 2002.

[3] CJ 4DPlex, <http://www.cj4dx.com/> (January 26, 2014).

[4] Y. Suzuki and M. Kobayashi, “Air Jet Driven Force Feedback in Virtual Reality,” IEEE Computer Graphics and Applications, vol. 25, no. 1, February 2005, pp. 44–47.

[5] M. Minakuchi, “Use of Wind for Information Acquisition,” Proceedings of IADIS International Conference/Interfaces and Human Computer Interaction 2007, July 2007, pp. 27–34.

[6] E. Sawada et al., “BYU-BYU-View: A Wind Communication Interface,” ACM SIGGRAPH 2007 Full-Conference DVD-ROM Emerging Technologies, August 2007.

[7] T. Moon and G. J. Kim, “Design and Evaluation of a Wind Display for Virtual Reality,” Proceedings of the ACM Symposium on Virtual Reality Software and Technology, November 2004, pp. 122–128.

[8] T. Kosaka, H. Miyashita, and S. Hattori, “Development and Evaluation of Immersive 3D Wind Display,” IPSJ Symposium Series, vol. 2007, no. 4, April 2007, pp. 105–112 (in Japanese).

[9] T. Nakano, S. Saji, and Y. Yanagida, “Indicating Wind Direction Using a Fan-Based Wind Display,” Proceedings of EuroHaptics 2012, vol. II (Lecture Notes in Computer Science, vol. 7283, Springer), Tampere, Finland, June 2012, pp. 97–102.

[10] Y. Yoshioka, T. Nakano, and Y. Yanagida, “Wind Direction Perception Using a Fan-based Wind Display: Effect of Head Position on the Discrimination Performance,” Proceedings of Joint Virtual Reality Conference 2012, Madrid, Spain, October 2012, pp. 25–26.

[11] M. L. Heilig, “Sensorama Simulator” US Patent 3,050,870, August 1962.

[12] S. Cardin, F. Vexo, and D. Thalman, “Head Mounted Wind,” Proceeding of the 20th Annual Conference on Computer Animation and Social Agents, June 2007, pp. 101–108.

[13] A. Lehmann, C. Geiger, B. Wöldecke, and J. Stöcklein, “Design and Evaluation of 3D Content with Wind Output,” Proceedings of IEEE 3D User Interfaces 2009, March 2009, pp. 14–15.

[14] H. Matsukura, T. Yoneda, and H. Ishida, “Smelling Screen: Development and Evaluation of an Olfactory Display System for Presenting a Virtual Odor Source,” IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 4, April 2013, pp. 606–615.

[15] K. Hirota, Y. Ito, T. Amemiya, and Y. Ikei, “Generation of Direction Wind by Colliding Airflows,” In Proc. of IEEE World Haptics Conference 2013, Korea, April 2013, pp. 509–514.

[16] H. Kubota, “Effect of Indoor Air Movement on Warmth and Comfortableness,” Heating, Air-Conditioning and Sanitary Engineers, vol. 58, no. 3, March 1984, pp. 243–248 (in Japanese).

[17] Y. Kojima, Y. Hashimoto, and H. Kajimoto, “A Novel Wearable Device to Present Localized Sensation of Wind,” Proceedings of ACM International Conference on Advances in Computer Entertainment Technology, October 2009, pp. 61–65.

[18] W. Hashimoto and R. Imaeda, “Tactile Perception of Artificial Breezes with Human Finger,” Correspondences on Human Interface, vol. 8, no. 3, 2006, pp. 35–40 (in Japanese).

[19] J. H. McDonald, Handbook of Biological Statistics. Baltimore, MD: Sparky House Publishing, 2009.

Usability Evaluation of Digital Games for Stroke Rehabilitation in Taiwan

Mei-Hsiang Chen
School of Occupational Therapy
Chung Shan Medical University/Chung Shan Medical University Hospital
Taichung 40201, Taiwan
cmh@csmu.edu.tw

Lan-Ling Huang and Chang-Franw Lee
Graduate School of Design
National Yunlin University of Science and Technology
Yunlin 64002, Taiwan
g9630806@yuntech.edu.tw
leecf@yuntech.edu.tw

Abstract—**This study aimed to compare the effectiveness and usability of using conventional devices, Wii and XaviX, in rehabilitation, and to propose advices to improve the digital game design. A clinical trial was implemented to evaluate the effectiveness and the usability of using Wii and XaviX in rehabilitation; they were assessed by stroke patients. Twenty-eight stroke patients were divided in three groups: Conventional, Wii, and XaviX groups. The results can be summarized into the following points: (1) The effectiveness index in each group indicated improvements for upper extremity function in stroke patients. However, the effectiveness indexes of each assessment between three groups do not have significant differences. (2) All patients in this clinical trial had fun when using the digital gaming devices for rehabilitation. (3) The suggestions for improving the design in digital games are as follows: to increase difficulty, and the response time, levels of the games need adjustment; to record movement data and game scores each time, the hand controller must be interchangeable for the users; the controller should be adjustable to fit different hand dimensions of the patients; the game and controller movements need to be designed to correspond to real-life activities; and the controller's operation needs to be simplified. In order to make these devices more suitable to use in rehabilitation, a comprehensive follow-up design development based on these proposed guidelines would be necessary in order to embody design improvements of the devices.**

Keywords- commercial video game; stroke; upper extremity rehabilitation.

I. INTRODUCTION

Stroke is one type of the cerebrovascular disease threatening people in modern societies. Upper-extremity motor deficit is one of the main symptoms for the stroke patients [1]. Many daily living tasks are performed by the upper extremity; therefore, rehabilitation treatment of upper extremity is very important for stroke patients. In order to restore the movement function of patient's upper extremity, patients must accept the rehabilitation activities. A rehabilitation device is an essential tool in the process of rehabilitation therapy. A study [2] has found that the current clinical upper extremity rehabilitation devices are mostly static and provide no feedback to the patients. Patients easily feel bored while repeating the same activity, hence generating a negative attitude toward the therapy process [3][4][5]. In order to increase mental satisfaction and physical vitality in rehabilitation therapy, some therapists have tried to use the off-the-shelf digital game devices in rehabilitation and have found effective treatment outcomes in in addition to enhancing the patient's treatment motivation [4][6][7].

There are already many studies focused on digital game devices in rehabilitation such as Wii [4][5][7][8][10], Playstation EyeToy [6][9] and Kinect [5]. For example, a study examined the feasibility and safety of the Wii gaming system and compared it with recreational therapy in facilitating motor function of the upper extremity [8]. The results showed that Wii gaming technology represents a safe, feasible, and potentially effective way to facilitate rehabilitation therapy and promote motor recovery after stroke. However, some of Wii's disadvantages were that stroke patients found the control of the handset buttons difficult and frustrating to use; however, this obligatory hand use improved gross motor dexterity. Stroke patients did not simply play Wii sports, rather the device was used as a rehabilitation tool with targeted and movement goals aimed at reinforcing appropriate and coordinated motor patterns [7].

The above-mentioned studies reveal that there are some usage problems to directly apply the off-the-shelf digital game devices in clinical rehabilitation. User-centered design emphasizes that designers observe the behaviors and usage situation of the actual user's experience, that is, how users use a product and what problems and needs exist. About the medical and rehabilitation equipment, variables for evaluation may include effectiveness, ease of use, comfort, and acceptability [10]. Through design, these off-the-shelf digital game devices can be tailored for different stroke patients for their respective requirements in the rehabilitation therapy. It will make the devices more suitable for rehabilitation therapy, and enhance the use safety and effectiveness for stroke patients.

These existing devices are originally designed for entertainment with normal people with healthy physical and action conditions, not intended for rehabilitation therapy purposes or for people with physical disabilities [6]. Further confirmation and evaluation is necessary to see if a digital game device can really meet the user's usability needs in addition to its rehabilitation effectiveness.

Nintendo Wii and XaviX have been tested in clinical rehabilitation in several hospitals in Taiwan. However, whether the superiority of Virtual Reality systems can facilitate conventional therapy currently in use remains to be determined [3][11][12]. Further development is required to ensure that these devices are easy for patients and therapists to set up in a clinic [6]. A pilot trial was implemented to evaluate the effectiveness, usability and satisfaction between conventional devices, Wii, and XaviX in rehabilitation [13]. However, the result of effectiveness showed no statistically significant improvements between groups, due to the number of patients is limited. Therefore, this report aimed to increase larger number of patients to compare the effectiveness and usability of using conventional devices, Wii and XaviX in rehabilitation, and to propose advices on improve the digital game design.

In the following sections, the methods applied and results obtained will be described respectively, and then the implication of the results will be discussed, followed by a brief conclusion.

II. METHODS

A clinical trial was implemented to evaluate the effectiveness between conventional devices, Wii, and XaviX in rehabilitation. In addition, the usability of using Wii and XaviX in rehabilitation were also assessed by stroke patients. The following two sections describe the research contents for these two kinds of subjects, respectively.

A. *Effectiveness evaluation*

Stroke patients were recruited from the occupational therapy department of Chung Shan Medical University Hospital. Inclusion criteria were the following: (a) Hemiparetic with upper extremity dysfunction following a single unilateral stroke, (b) a history of first-time stroke (3-48 months post stroke), (c) the required upper extremity rehabilitation convalescent levels were Brunnstrom stage III to IV, i.e., having basic upper extremity synergies to perform joint movement voluntarily, (d) ability to communicate, (e) able to understand and follow instructions. Exclusion criteria were the following: (a) engaged in any other rehabilitation program during the study and (b) serious aphasia or cognitive impairment. Each patient gave informed consent. This study was approved by the Human Research Ethics Board of Chung Shan Medical University Hospital.

A total of 28 post-stroke patients were admitted to the Occupational Therapy Department of Chung Shan Medical University Hospital. The mean age of the patients was 53.1 years (SD 15.1), with a mean post-stroke time of 9.2 months (SD 7.5). The characteristics of the patients in the three study groups are shown in Table 2. As the data were not normally distributed, non-parametric tests were conducted. There were no statistically significant differences among the three groups with regard to age, time since stroke onset. Of the 4 who did not finish, 1 was transferred out of the hospital, 1 was discharged, and the other 2 failed to continue treatment. Twenty-four consecutively screened stroke patients finally completed the trial.

Conventional equipment, Wii, and XaviX, were used in this trial (Table 1). Each group was assigned to use two games or equipment in the additional treatment. The games and equipment for the groups were selected by three occupational therapists, and were considered as similar in training effect on upper extremity movements. For Nintendo Wii, two games (boxing and bowling) of Wii Sport were selected to use in this trial. As for XaviX, two games (bowling and ladder climbing) were selected to use in this trial. Ladder climbing contains three levels (easy, normal and difficult). While playing bowling, the user needed to hold a soft bowling ball, fixed to the hand with a safety belt. In ladder climbing, the user needed to wear glove sensors in the palms of both hands [14]. Corresponding to the Wii games (Bowling and Boxing) and the XaviX games (Bowling and Ladder climbing), two conventional equipment (Curamotion exerciser and the Climbing board and bar) were selected in this trial. Both are commonly used in rehabilitation.

Four functional assessments were used as follows:

a) Fugl-Meyer Assessment of Physical Performance (FMA) [15]. It was used to evaluate the motor functions. The upper extremity motor test part with a possible highest score of 66 was adopted in the evaluation. The reliability of Fugl-Meyer Assessment is generally considered reliable [16].

b) Box and Block Test of Manual Dexterity (BBT). It was used to test gross manual dexterity of a patient's affected side [17]. In the test, the patient was asked to move as many cubes (of side length of 2.5 cm) as possible using only the thumb and index fingers during a timed 60s trial.

c) The Functional Independence Measure (FIM). This scale assesses physical and cognitive disability. The scale includes 18 items, of which 13 items are physical domains based on the Barthel Index and 5 items are cognition items. Each item is scored from 1 to 7 based on level of independence, where 1 represents total dependence and 7 indicates complete independence. This measurement was assessed and shown to have high reliability and validity [18][19][20].

d) Upper extremity range-of-motion (ROM). This is used to assess the passive and active range-of-motion on the affected side [21].

The training comprised 20 sessions during 2 months, with each session lasting 30 minutes (excluding set-up time). The effectiveness was evaluated before and after completing the 20 training sessions. In addition to these trainings (Wii, XaviX, and Conventional) in this study, all patients also received at least 1 hour of occupational therapy and 1 hour of physical therapy. After training, all patients asked to fill in a questionnaire. The content of the questionnaire is shown in the following section.

B. *Usability evaluation by the patients*

Three 5-point Likert type questions about satisfaction, motivation, and fun were then presented on a sheet of paper for the patients to answer. By this trial, we expected to understand the patients' satisfaction with using the digital gaming devices in clinics.

C. *Usability evaluation by the occupational therapists*

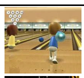
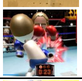
The method of expert interview was conducted to survey the current situation of the use of digital game devices and evaluate the usability of such devices in upper extremity rehabilitation therapy. The locations, participants, contents of the interview and process are as follows.

The selection criteria for the interview subjects were mainly based on hospital size, type (public vs. private), geographic location, and types of digital games used. Two hospitals were selected in a preliminary investigation: Taipei Veterans General Hospital (TVGH) and Kaohsiung Veterans General Hospital (KVGH). Wii was used in Kaohsiung Veterans General Hospital and XaviX was used in Taipei Veterans General Hospital.

The occupational therapist plays a major role in the rehabilitation process to determine how the digital game device is used. As a professional, the occupational therapist possesses expertise about therapy theory and experiences, which are useful for the evaluation of rehabilitation devices or the commercial products applied in rehabilitation.

Selection criteria for the interviewed therapists were: at least 5 years of work experience in occupation therapy, and at least one year experience in adopting the digital game device intervention in rehabilitation treatment. In total, eight therapists were selected and interviewed, three (one male [A] and two females [B, C]) from KVGH and five (two males [D, E] and three females [F, G, H]) from TVGH. They have an average age of 35.1 yrs (SD=6.3) and work experience of 10.8 years (SD=6.3).

TABLE I. GAMES AND CONTROLLERS OF NINTENDO WII AND XAVIX

Groups	game		controller	usage situation
Nintendo Wii	bowling			
	boxing			
XaviX	bowling			
	ladder climbing			

Semi-structured interviews were used, which mainly consisted of two parts: (1) Therapist personal profile - therapist gender, age, hospital name, and work experience in years, and (2) Questions about the usability evaluation of existing digital gaming device: The questions were:

1. Effectiveness: How do you feel about the therapy effectiveness of the device in restoring the patient's upper extremity functions?
2. Ease of use: Do you have any usage problems in setting up the device to be used for therapy? What needs to be improved in the design of the devices?
3. Acceptability: How do the patients respond to the use of digital game device in rehabilitation? Do they accept it and like it?

D. *Data analysis*

All data were analyzed with SPSS for Windows version 13.0 (IBM SPSS, Inc., Chicago, IL, USA). Intention-to-treat analysis (ITT) was used to analyze all data in each group [22]. The missing subjects who actually did the pre-test were included as subjects who had received treatment, and their pre-test scores were addressed so that they equaled their post-test scores; however, these scores indicated very little about the efficacy of the treatment. The characteristics of the study groups were described as mean and SD. The Wilcoxon signed rank test was the non-parametric alternative to the two-related samples tested for pre- and post-tests in each group, and the Kruskal-Wallis one way analysis of variance by ranks test was used for differences in the effectiveness index between groups. Differences were considered significant when $p<0.05$. For each group, the effectiveness index of each functional assessment was calculated, i.e., the post-test score minus the pre-test score, and then divided by the maximum possible progress (possible highest score of the assessment minus pre-test score) [23]. To analyze the interview data from occupational therapists, the recording was firstly transcribed verbatim. Similar opinions were combined and all unique responses were independently itemized for further discussion. To analyze the questionnaire, data from stroke patients who completed treatment according to the devices used (conventional devices, the Wii device and the XaviX device). For each question on the questionnaire, the mean and standard deviation were calculated.

III. RESULTS AND DISCUSSION

A. *Within-group differences in score changes for each group*

Comparison of the pre- and post-test subjects in each group showed significant improvements for upper extremity function in stroke patients (Table 3). The XaviX group had significant differences on four assessments: the FMA (p=0.009), BBT (p=0.022), FIM (p=0.009), ROM-proximal (p<0.001), and ROM-distal (p<0.001). However, the conventional group and the Wii group had significant differences on three assessments: the FMA, FIM, and ROM. In addition, the results showed that the effectiveness index of each group indicated improvements for upper extremity function in stroke patients on the FMA, BBT, FIM, and ROM (Table 3).

B. *Between-group differences in score changes of effectiveness index*

The Krusal-Wallis one-way analysis of variance ranks showed no significant differences among the three groups for the FMA (p=0.349), BBT (p=0.950), FIM (p=0.644), ROM-proximal (p=0.186), and ROM-distal (p=0.114). On the other hand, comparing the effectiveness index of each assessment in three groups, we found that the effectiveness index of the Wii group on the FMA (mean=0.37, SD=0.21) indicated greater improvement than that of the XaviX group (mean=0.27, SD=0.33) and conventional group (mean=0.26, SD=0.05). The mean score of the effectiveness index of the XaviX group was close to that of the conventional group.

TABLE II. CHARACTERISTICS OF STROKE PATIENTS

Characteristics		Conventional	Wii	Xavix	P-value*
Number of subjects		8	9	11	
Gender (male/female)		5/3	6/3	7/4	
Age (years)	mean (SD)	48.5 (16.4)	49.6 (15.1)	59.5 (13.1)	0.184
Time from stroke onset to involvement in gaming (months)	mean (SD)	6.3 (3.7)	9.8 (3.5)	10.7 (11.0)	0.143
Paretic side (left/right)		4/4	4/5	5/6	
Brunnstrom stage	proximal (median, range)	4.0 (4-4)	4.0 (4-6)	4.0 (3-5)	
	distal (median, range)	4.0 (2-4)	4.0 (3-6)	4.0 (1-5)	

*P for conventional group versus Wii group versus XaviX group
FMA (UE): The upper extremity portions of the motor subscale of the Fugl-Meyer Assessment; FIM: Functional Independence Measure; BBT: Box and Block Test of Manual Dexterity; ROM: Upper extremity range-of-motion.

TABLE III. WITHIN-GROUP DIFFERENCES IN CHANGE SCORES FOR PRE- AND POST-TEST IN EACH GROUP

assessments	Conventional		Wii		Xavix	
	mean (SD)	P-value [a]	mean (SD)	P-value [a]	mean (SD)	P-value [a]
FMA (UE) (pre-test)	27.0 (7.0)	0.006**	42.2 (13.9)	0.006**	39.6 (18.6)	0.009**
FMA (UE) (post-test)	37.9 (4.5)		50.3 (10.5)		44.7 (18.4)	
BBT (pre-test)	8.5 (10.5)	0.103	23.9 (16.6)	0.070	24.5 (20.4)	0.022*
BBT (post-test)	10.0 (12.0)		28.2 (18.5)		26.7 (21.8)	
FIM (pre-test)	72.3 (12.8)	0.009*	80.3 (9.1)	0.021*	75.5 (9.1)	0.009**
FIM (post-test)	96.6 (11.1)		82.6 (7.8)		79.1 (7.5)	
ROM-proximal (pre-test)	63.9 (41.7)	0.000**	81.8 (48.5)	0.000**	76.5 (46.7)	0.000**
ROM-proximal (post-test)	80.1 (47.6)		91.1 (50.7)		88.8 (45.8)	
ROM-distal (pre-test)	18.1 (27.4)	0.000**	41.6 (33.9)	0.000**	30.6 (36.0)	0.000**
ROM-distal (post-test)	27.3 (29.8)		50.0 (32.9)		39.0 (37.6)	

a P for Within-group differences in change scores for pre- and post-test
* Significant at ≦0.05 level
** Significant at ≦0.01 level

C. *Results of usability assessment by stroke patients*

Motivation. Did it promote your treatment motivation when you used the Wii, XaviX or conventional equipment for treatment? Patients of the Wii and XaviX groups agreed that digital games used in the treatment promoted their treatment motivation (Wii: mean=3.8, SD=1.0; XaviX: mean=3.8, SD=0.9; Conventional: mean=3.5, SD=0.5).

Fun. Do you feel that using the Wii, XaviX or conventional equipment for treatment was more interesting than traditional rehabilitation devices? Patients of the Wii and XaviX groups agreed that these devices were more interesting than traditional rehabilitation devices (Wii: mean=4.3, SD=0.9; XaviX: mean=4.4, SD=0.5; Conventional: mean=2.3, SD=0.9). Seven of them specially mentioned that these devices provided useful information, such as the scores, audio and video feedbacks, as well as interesting interactions, making them feel positively toward doing the treatment activities.

D. *Results of usability assessment by occupational therapists*

Effectiveness: How do you feel about the therapy effectiveness of the devices in restoring the patient's upper extremity functions? For effectiveness of Wii, all three therapists with experience in Wii (A, B, and C) agreed that the Wii Sport was effective in upper extremity rehabilitation, in addition to enhancing the patient's treatment motivation. In addition, Therapist A suggested that Wii Fit was especially suitable for balance therapy for stroke patients. For effectiveness of XaviX, all five therapists with experience in XaviX (D, E, F, G, H) agreed that the device is effective in upper extremity rehabilitation, and could enhance the patient's treatment motivation and pleasure. D also commented that some of the game projects with their controllers can effectively strengthen the training of reaching and grasping movements and achieve better therapy effectiveness as compared to the traditional devices. For instance, when the patient wears the gloves as controllers in the XaviX games, his/her hands are required to actually do the reach and grasp movements to complete the game task. In so doing, the training of the hands in those movements are apparent and practical. In addition, therapists E and F mentioned that the existing traditional devices are still important in rehabilitation and not possible to be fully replaced by digital gaming devices. The traditional devices do not require the patient to react in time and restrict his/her movement speed, hence the patient can repeat the movement as many times as he/she wishes and at a speed at his/her own control. Therefore, the digital game device can play a supporting role in rehabilitation treatment by providing diversification and interesting game projects for improvement of patient motivation toward the treatment.

Ease of Use: Do you have any usage problems in setting up the devices to be used for therapy? What needs to be improved in the design of the devices? For the Wii, all three therapists (A, B, C) considered that the device is easy to set up, except some errors may occur when setting the game software items. The usage problems of this device are:

- The current software interface is in Japanese and not easy to understand, hence is prone to cause errors in the set-up process;
- The required response time of the game is too fast and not easy to keep up with by the patients;
- Some patients may have difficulty to hold the hand controller, hence need additional bandage to tie it on the hand;

- For some patients, the games are too difficult.

For XaviX, all five therapists agreed that the hardware is easy to set up, except the software interface operation may cause occasional mistakes. The usage problems of this device are:

- The current software interface is in Japanese and not easy to understand, hence is prone to cause errors in the set-up process, especially for the games of visual perception and memory training, which are impossible to operate without literacy of the Japanese language;
- The sensor is not sensitive enough, e.g., the action may obstruct the reflective film and interrupt the detection by the sensor;
- The controller gloves have only one single size, do not necessarily fit to all patients;
- Although the games are available in three levels of difficulty, the differences between the difficulties are too abrupt and hard to meet the patient's required degree of difficulty.

Acceptability: How do the patients response to the use of digital game devices in rehabilitation? Do they accept it and like it? The result shows that all eight therapists agreed that most patients can accept the use of Wii or XaviX for treatment. Therapist B explained that people interested in the game have a very high degree of acceptance of such devices. However, some elderly patients tend to prefer using the traditional rehabilitation devices, because they feel that the traditional devices are some physical objects that can be held and manipulated, hence giving a feeling of the effect of treatment. On the contrary, digital game devices give an impression of laxness and frivolity, and hence are inefficient and ineffective.

IV. CONCLUSION

This study conducted a clinical trial to examine the therapeutic effectiveness and usability of using conventional devices, Wii and XaviX on upper extremity functional recovery after stroke. We also used the intention-to-treat concept to analyze all data (except for motivation and enjoyment) in order to avoid overoptimistic estimates of the efficacy of an intervention resulting from the removal of non-compliers by accepting noncompliance and protocol deviations. The effectiveness index in each group indicated improvements for upper extremity function in stroke patients. However, the effectiveness indexes of each assessment between three groups are not significant differences.

The results obtained from the interviews, it was found that the digital game devices currently used in clinical rehabilitation have been evaluated by professional therapists as useful and effective in supporting the treatment, though some design improvements may be necessary. In order to involve stroke patients in rehabilitation treatment as users of game devices for general players, factors concerning usability and safety in terms of interface with the stroke patients need to be further enhanced. Design guidelines concerning the improvement of existing digital game devices can be synthesized as follows, where items a to d are about software design, and items f to i about hardware: (a) To increase the response time of the games. (b) To increase difficulty levels of the games in order to better suit the various patients with different abilities of upper extremity functions. (c) To expand the sensor's sensing scope. (d) To be able to record movement data, such as: reaction time, operating time. (e) To improve the ways to fix the controller on the user's hand. (f) To fit the controllers size for different hand dimensions of the patients. (g) To provide better correspondence between the game and real-life movements. (h) To provide controllers for body control training, such as chest strap and belt. (i) To simplify the controller's operation.

Two suggestions are proposed for future studies: (1) the therapists in the study had all chosen to use digital devices for intervention for at least a year, it may biased the results. It will be better if the views of those therapists who choose not to use digital games were also further surveyed. (2) In order to make these devices more suitable to use in rehabilitation, a comprehensive follow-up design development based on these proposed guidelines would be necessary in order to embody design improvements of the devices.

ACKNOWLEDGMENT

This study is supported by the National Science Council of the Republic of China with grant No: NSC 100-2221-E-040-009-MY2.

REFERENCES

[1] C. Gowland, H. DeBruln, J. Basmajian, N. Piews, I. Burcea, "Agonist and antagonist activity during voluntary upper-limb movement in patients with stroke," Physical Therapy, vol. 72, pp. 624-633, Sep. 1992, [Online]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1508970

[2] L. Huang, C. Lee, M..Chen, "The investigation of design improvement requirements for the upper extremity rehabilitation devices in Taiwan," The 11th Asia Pacific Industrial Engineering and Management Systems Conference (APIEMS 2010), Dec. 2010, pp.1-6.

[3] G. Saposnik, M. Levin, stroke outcome research Canada working group, "Virtual Reality in Stroke Rehabilitation: A meta-analysis and implications for clinicians," Stroke, vol.42, pp.1380-1386, May. 2011, doi: 10.1161/STROKEAHA.110.605451. Epub 2011 Apr 7.

[4] L. Joo et al., "A feasibility study using interactive commercial off the shelf computer gaming in upper limb rehabilitation in patients after stroke," Journal of Rehabilitation Medicine, vol.42, pp.437-441, May. 2010, doi: 10.2340/16501977-0528.

[5] Y. Chang, S. Chen, J. Huang, "A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities," Research in Developmental Disabilities, vol.32, pp.2566-70, Nov–Dec. 2011, [Online]. Available from: http://www.sciencedirect.com/science/article/pii/S0891422211002587

[6] B. Lange, S. Flynn, A. Rizzo, "Initial usability assessment of off-the-shelf video game consoles for clinical game-based motor rehabilitation," Physical Therapy Reviews, vol.14, pp.355-363, 2009, doi: 10.1179/108331909X12488667117258

[7] M. Mouawad, C. Doust, M. Max, P. McNulty, "Wii -based movement therapy to promote improved upper extremity function post-stroke: a pilot study," Journal of Rehabilitation

Medicine, vol. 43, pp.527–533, May. 2011, doi: 10.2340/16501977-0816.

[8] G. Saposnik et al. “Effectiveness of virtual reality using Wii gaming technology in stroke rehabilitation: a pilot randomized clinical trail and proof of principle,” Stroke, vol.41, pp.1477-1484, Jul. 2010, doi: 10.1161/STROKEAHA.110.584979.

[9] G. Yavuzer, A. Senel, M. Atay, H. Stam, "Playstation eyetoy games" improve upper extremity-related motor functioning in subacute stroke: a randomized controlled clinical trial,” vol.44, pp.237-44, Sep. 2008, [Online]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18469735

[10] K. Jacobs, (3rd ed). “Ergonomics for therapists”. Mosby Elsevier. Louis, Missouri 63146, 2008.

[11] P. Langhorne, F. Coupar, A. Pollock, “Motor recovery after stroke: a systematic review,”Lancet Neurology, vol.8, pp.741-754, Aug. 2009, doi: 10.1016/S1474-4422(09)70150-4.

[12] L. Lucca, “Virtual reality and motor rehabilitation of upper limb after stroke: a generation of process?,”Journal of Rehabilitation Medicine, vol.41, pp.1003-1100, Nov. 2009, doi: 10.2340/16501977-0405.

[13] L. Huang, C. Lee, M. Chen, “Applying commercial digital games to promote upper extremity movement functions for Stroke patients,” The Sixth International Conference on Advances in Computer-Human Interactions (ACHI 2013) IARIA, Feb. 2013, pp.1-6, ISSN: 2308-4138, ISBN: 978-1-61208-250-9

[14] XaviX Port. User’s guide. retrieved: 02,2014. http://www.xavixstore.com/products/users_guide/port_ug.pdf 2014.02.15.

[15] A. Fugl-Meyer, L. Jaasko, I. Leyman, S. Olsson, S. Steglind, “The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance,” Scandinavian Journal of Rehabilitation Medicine, vol.7, pp.13-31, 1975, [Online]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1135616

[16] P. Duncan, M. Propst, S. Nelson, “Reliability of the Fugl-Meyer assessment of sensorimotor of sensorimotor recovery following cerebrovascular accident,” Physical Therapy, vol.63, pp.1606-1610, Oct. 1983, [Online]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6622535

[17] V. Mathiowetz, G. Volland, N. Kashman, K. Weber, “Adult norms for the Box and Block Test of manual dexterity,” The American Journal of Occupational Therapy, vol.39, pp.386-391, Jun. 1985, [Online]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3160243.

[18] L. Brosseau and C. Wolfson, “The inter-rater reliability and construct validity of the Functional Independent Measure for multiple sclerosis subjects,” Clinical Rehabilitation, vol.8, pp.30-41, May. 1994, doi: 10.1177/026921559400800203

[19] K. Ottenbacher, Y. Hsu, C. Granger, R. Fiedler, “The reliability of the functional independent measure: a quantitative review,” Archives of Physical Medicine and Rehabilitation, vol.77, pp.1226-1232, Dec. 1996, [Online]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8976303

[20] J. Ravaud, M. Delcey, A. Yelnik, “Construct validity of the functional independent measure (FIM): questioning the unidimensionality of the scale and the “value” of FIM scores,” Scandinavian Journal of Rehabilitation Medicine, vol. 31, pp.31-41, Mar. 1999, [Online]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10230001

[21] C. Heck, I. Hendryson, C. Rowe, “Joint motion: method of measuring and recording,”American Association of Orthopaedic Surgeons, 1965.

[22] S. Gupta, “Intention-to-treat concept: A review,”Perspectives in clinical research, vol. 2, pp.109-112, Jul. 2011, doi: 10.4103/2229-3485.83221.

[23] C. Lee, C. Wu, I. Lien, M. Hsu, K. Lin, “Effects of modified constraint-induced movement therapy on stroke patients,”Formosan Journal of Medicine, vol. 10, pp.429-437, 2006, [Online]. Available from: http://www.ntuh.gov.tw/PMR/Lists/List14/Attachments/10/11999587481.pdf

BOrEScOPE – Exoskeleton for Active Surgeon Support during Orthopedic Surgery

P. P. Pott, M. Hessinger, R. Werthschützky,
H. F. Schlaak
Institute of Electromechancial Design
Technische Universität Darmstadt
Darmstadt, Germany
p.pott@emk.tu-darmstadt.de

E. Nordheimer, E. Badreddin,
A. Wagner
Institute of Computer Engineering, Automation Lab
Heidelberg University
Heidelberg, Germany
achim.wagner@ziti.uni-heidelberg.de

***Abstract*— The use of robots in a medical environment is a challenging task not only for system development but also for the actual application in this demanding environment. Robots are used to enhance surgery quality in terms of precision, application of new therapies, or to improve ergonomics - only to name a few reasons. The approach described in this paper is to provide a lightweight exoskeleton worn by the orthopedic surgeon. It is intended to be used during drilling tasks at the spine and to enhance precision as the surgeon is led by optic, acoustic, and haptic perception. The parallel flux of forces and the inherently mobile robot base allow the surgeon to directly maintain responsibility for surgery. Not only the mechanical design of the system but also the control is decomposed into several levels. To do so, a behavior-based approach is used. The system's design criteria are briefly described and first results are presented. The exoskeleton is composed of an anthropomorphic arm actuated by twisted-string actuators. This leads to a lightweight construction. To provide sufficiently fast and precise information about the spatial position and its time derivations, optical and inertial tracking is used. A User Guidance Opto-Acoustic Display is utilized to provide the surgeon with information on position and orientation of the tool in six degrees of freedom with respect to the desired trajectory. First experimental results derived that the intended workspace meets the surgical requirements and the user guidance system enables the surgeon to follow the desired trajectory by intuitive user guidance.**

Keywords- exoskeleton; orthopedic surgery, human-machine interaction; twisted-string actuation; behavior-based system decomposition

I. INTRODUCTION

Medical robotic systems for the use in the operating room (OR) have been under development for more than 20 years. Early systems for neurosurgery [1, 2] and orthopedics [3] proved usefulness and even made it for commercialization. However, their impact was not as high as expected [4]. In the last ten years, many new robotic systems have been developed and even introduced to the market. The most popular is the daVinci Surgical System by Intuitive Surgical, Inc., Sunnyvale, CA, USA. Nevertheless, there are hundreds of different systems and many reviews to learn more about the field of robotics.

Aim of our work is to develop and to design a robotic interaction system for orthopaedic surgery. Here, the surgeon has to fulfil delicate tasks like drilling the spine while maintaining high precision in the sub-millimeter range. Placing a robotic arm next to the OR table [5], the ceiling [6] or even on the patient [7] does not seem to be appropriate. Earlier work of our group showed the high potential of placing the robot in the user's hand [8, 9] to compensate tremor and involuntary movements both from surgeon and patient [10]. This robotic system provides precise movement and ease-of-use. However, its size and weight is not appropriate for longer deployment. Instead of using a passive balancing system we decided to develop a new system worn at the surgeon's arm near to his or her centre of gravity to improve ergonomic handling. In the following sections, we will present and describe the system's concept, basic components, the control strategy, and first results.

II. SYSTEM DESIGN

In this paper, we provide an overview of the BOrEScOPE system. It comprises an external optical high-speed tracking system for six degrees of freedom (DOF) position and orientation measurement fused with data from an inertial measurement unit (IMU), the robotic system including actuation and sensor systems and the mechanical part, the control hard- and software, and finally an opto-acoustic display unit for communication and user guidance. In the following sections, we will address these sub-systems and describe the control strategy.

A. Robotic System

The robotic system of the BOrEScOPE basically consists of an exoskeleton for the (right) arm of the surgeon including shoulder and wrist (Figure 1). All together seven degrees of freedom (DOF) are realized to provide good compliance with the human anatomy and the same dexterity. The range of motion of the shoulder (170° abd./add.; 150° flex./ex.; 180° inw./outw. rotation), elbow (100° flex.), and wrist (150° pron./sup.; 20° ulnar flex./ex.; 120° flex./ex.) joints has been derived experimentally. Shoulder elevation is not considered as the abduction angle is reduced to 80°. The arm is attached to a backpack that is carried by shoulder and hip harness.

To achieve a lightweight mechanism, the actuators are placed in the backpack and force is transmitted via Bowden cables. The actuators are based on the twisted string-concept [11], using a bunch of at least two strings that are drilled axially by a DC motor. This causes the string-arrangement to shorten and produces a rather high force. Using a lightweight ∅17 mm DC motor (1741 024 CXR by Faulhaber,

 ISBN: 978-1-61208-325-4

Schönaich, Germany) with 8 mNm nominal torque and three strings, a force of 130 N can be produced. Also, no traditional gear reduction is needed leading to very quiet operation. As only pulling forces can be produced, an antagonistic arrangement is used. Sensors are deployed at the string actuator to measure shortening and at the actuated joint to provide precise angle information. Doing so, the elasticity of the Bowden cable is used to derive a series-elastic actuator (SEA) [12]. Prior work of our group showed good results using SEA in human machine interaction [13]. The inherent compliance allows zero-torque control and robust reaction to dynamic external forces. This reduced stiffness "feels better" than a conventional robotic arm.

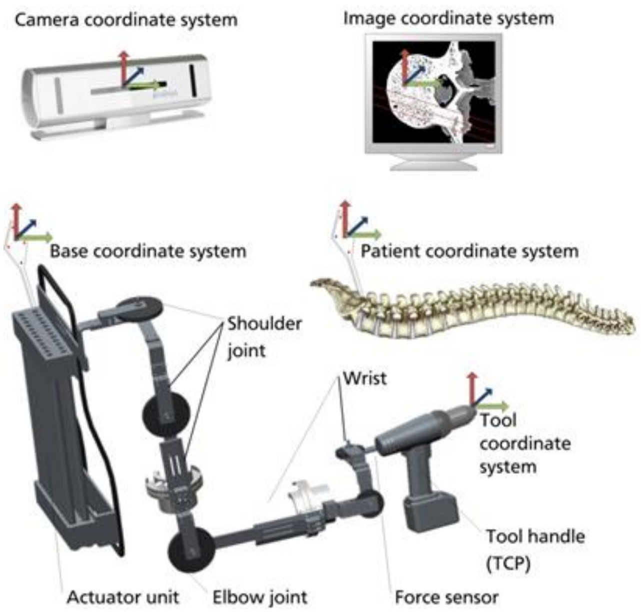


Figure 1 Overview of the robotic subsystem of the BOrEScOPE.

The system is designed to carry a 2.5 kg payload and compensate the gravity force of the human arm up to a body weight of 80 kg. Shoulder and elbow joints can provide speeds up to 6 rad/s. The static force to guide the user can be up to 10 N at the handle.

B. *Opto-Acoustic Display*

One of the challenges in developing a user-friendly Graphical User Interface (GUI) for the Human Machine Interaction (HMI) is to facilitate an intuitive operation and control of the technical system. The basic requirements are to reduce the possible error occurring during user interaction with the machine and to navigate the user. Since the tool position is influenced by the human tremor (frequency range of several Hertz), and since latencies in the feedback-loop must be avoided, a dynamical tool tracking is proposed, consisting of a combination of optical and inertial motion measurements. Based on these data the User Guidance Opto-Acoustic Display (UGOAD) is realized, which navigates the user to the goal pose (position and orientation), displays the processing trajectory, and gives a feedback of pose errors. Display and measurement latency has to be kept low to reduce phase shift in the feedback loop and to provide stable overall system behavior. The goal 6D poses as well as the processing trajectories are provided from planning data, which are defined by the surgeon using 3D patient imaging (CT). According to the requirements, a first UGOAD functional prototype was realized (see Figure 2).

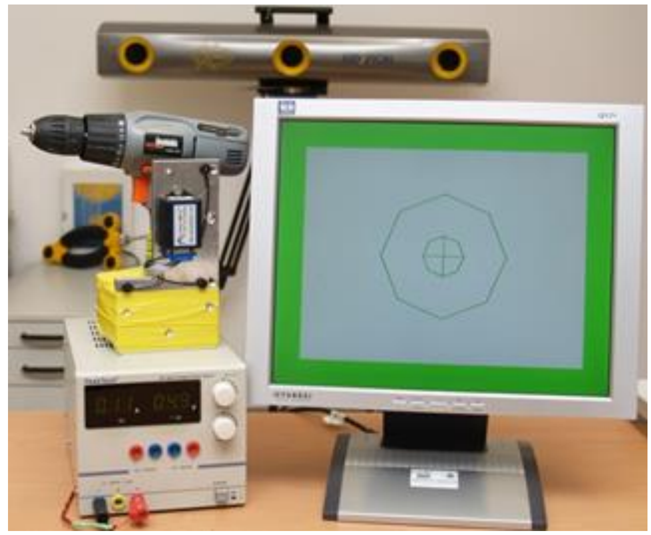


Figure 2 Experimental environment for the first prototype of the opto-acoustic display deployed in the BOrEScOPE system.

The experimental handheld drilling tool (Figure 2, on the left) was equipped with three active optical LED markers and an IMU device (Crista IMU, Cloud Cap Technology, Inc.). The monitor (Figure 2, on the right) can provide both optical and acoustic information. In the final implementation, a miniaturized screen will be attached directly to the tool. The optical tracking system (Krypton K600, Nikon Metrology, Inc.) (Figure 2, in the background) is used in addition to the Crista IMU to collect the motion data of the handheld device. Data fusion is accomplished using Kalman-filter based methods [14]. The resulting filtered variables for position, orientation, velocities, angular rates, and linear acceleration are utilized for navigation purposes and provided to the lower levels. In later development stages the complete handheld device can be mounted and aligned to the exoskeleton. The 6 DOF user navigation is realized by 2D representations of the tool pose on the UGOAD which is described below in detail.

C. *Control Structure*

The control system is developed according to Nested Recursive Behavior-based Control (RNBC) structure [15]. Accordingly, the hardware is realized as a number of components (Figure 3) interacting on diverse behavioral levels. In contrast to a one-to-one mapping of the behavior levels, one single behavior level can be distributed on multiple hardware components. Several behavior levels may be aggregated in one single hardware device. In the latter case, behaviors are realized as software processes. For the BOrEScOPE realization, the upper levels, i.e., *mission*, *navigation* and *trajectory control*, are realized as software processes integrated into a QNX-based (QNX Software Systems Ltd.) real-time PC. The behavior levels for *position control*, *collision avoidance*, *velocity control* and *force control* are realized using an embedded PC based on xPC Target™ (The Mathworks, Inc.). The xPC™ Target PC is interconnected with the QNX PC via a serial link and to the motor controllers (type EL7342 by Beckhoff Automation, Verl, Germany) via EtherCAT. The motor controllers directly control the currents of the actuators. Position constraints for link actuation are calculated using the robot kinematics in order to avoid internal collisions. Additionally, external ultrasonic (US) sensors can help to avoid collisions of the robot with its environment. A milling tool can be aligned with the patient coordinate frame and a target bone can be processed with the preplanned trajectory.

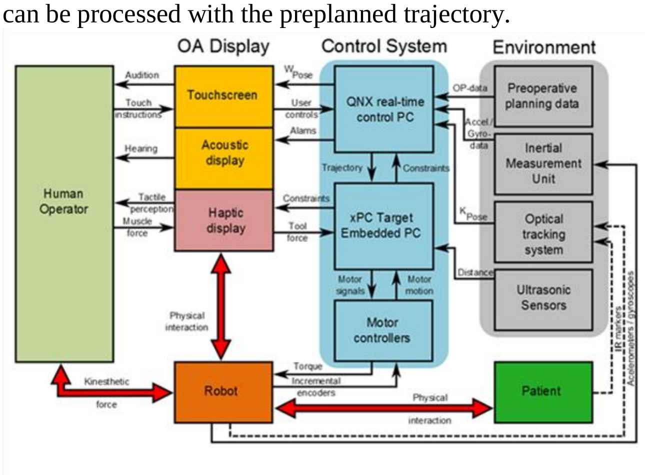


Figure 3 System architecture of the BOrEScOPE system

To achieve compliance with the behavior of the operator, three interaction modalities are realized: The opto-acoustic display provides optical (1) and acoustical output (2) while the robot provides haptic feedback (3). The control algorithm's input is a virtual static force field generated around the main axis of the bore and depending on the actual distance, speed and direction of movement of the BOrEScOPE's end effector [16, 17]. When the patient is moving, this force field also moves in space. To achieve smooth and comfortable movement, the real force acting between BOrEScOPE and the wrist are measured. The user tries to minimize the forces following the BOrEScOPE system.

Using this algorithm, the 7 DOF redundant robotic system can be controlled easily and intuitively while maintaining the human's dexterity. As both, the linear displacement at the actuators and the angular displacement in the actual joints are measured and controlled, serial-elastic actuation is achieved.

III. RESULTS

The BOrEScOPE system is still under development. The two main subsystems *opto-acoustic display* and *robotic system* show first and promising results that are described in the following.

The measured peak response time of hand movement as a result of optical stimuli amounts to around 250 ms. The requirement of visually provided information should be adapted on this process time. The reaction time of the UGOAD as well as the robot must be kept within a limit of 10-20 ms (10-20-fold faster). Thus, the calculation of graphical contents and of the control algorithm should terminate within this time. Based on this knowledge the sensor data acquisition, the global-control loop, and UGOAD were implemented as real-time processes in the QNX Neutrino operating system.

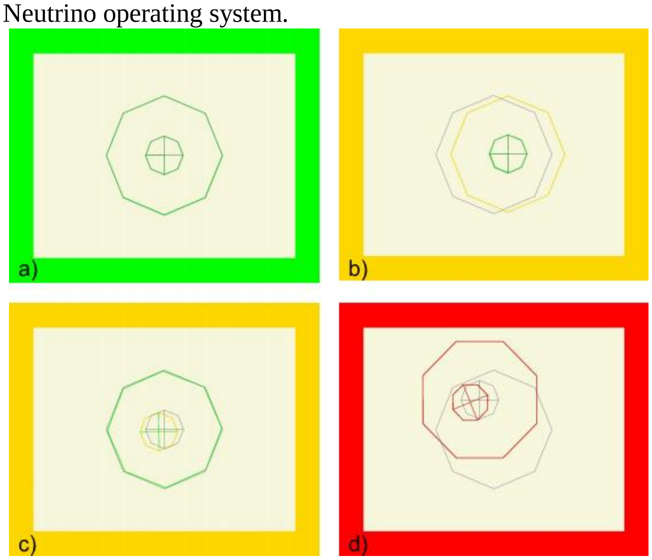


Figure 4 Three operate zones of the UGOAD: a) No displacement, b) Translational displacement in x-axis, c) Rotational displacement around x- and y-axis, d) Displacement in view axes

The first display prototype was realized by a 2D representation of the 6 DOF pose data. Accordingly, the actual and the reference pose of the tool are shown in the x-y-plane of the display. The z-axis is perpendicular to the display plane. In order to intuitively capture the 6 DOF contents in the 2D image a two body projection metaphor is realized. In this imagination one small colored octagon is mounted virtually at the tool tip and one large colored octagon at the rear of the tooling machine. Looking from above in direction of the drilling tool (z-axis) corresponds to looking through the large octagon and through the small octagon on the tool tip, which is in the center of both. The small black octagon with crosshairs and large black octagon are virtually mounted at the target (reference) pose. If the tool is aligned (Figure 4a), the small octagons are aligned and the large colored octagon has its original size in the central position. If the tool is misaligned in the x-axis (Figure 4b) the large colored octagon is shifted correspondingly in the x-direction. The same holds for the y-axis. A misalignment in the z-axis is represented by the size of the large colored octagon. A deviation in the positive z-direction means that the tool is too far away from the user, which is shown by the reduced size relatively to the large black octagon. Negative deviation means, that the tool is too narrow, displayed as increased size. A deviation in the orientation is displayed as shift of the small colored octagon.

For example, if the tool is turned around the y-axis (Figure 4c), the tool tip is moved in x-direction, displayed as x-axis-shift of the small colored octagon. The corresponding principle holds for the orientation error around the x-axis. Here, a y-shift of the small octagon can be observed. The orientation error around the z-axis is directly displayed as a rotation of the colored octagons around their centers.

As additional element, a rectangular border is shown in green color, which indicates that the pose is in the desired workspace. If the tool approximates the limit positions for at least one axis, the color changes firstly from green to orange, showing that a user intervention is required. In critical vicinity to the constraints the color changes to red (Figure 4d) asking for urgent motion actions. The color change is supported by changing the waveform of the acoustic channel.

IV. Discussion and Conclusion

To set up a robotic system with close human-machine interaction in a medical environment is a delicate task. However, the project is still in progress and work starting from the presented concept to the final realization is still ongoing. We managed to define interfaces between the robotic system and the human operator not only mechanically but also visually and using the audio channel. Smooth and comfortable working with the system is strongly dependent on low latency, high update rates, and actually predictable behavior. Here, our system will have to deal with some drawbacks as the force field generation is depending on data quality of the optical tracking system which tends to jitter and noisy signals. This will be addressed in future by using redundant LED markers and by combining data of an inertial tacking system. Furthermore, the quality of real-time data transfer will be improved.

First tests with users demonstrate that the 6 DOF can be captured by the majority of subjects without further explanation. Thus, the usage of the UGOAD as a feedback in the human interaction with the machine implicates a massive improvement of human performance to achieve the common tasks and there is every indication that the developed UGOAD insure an intuitive operation and an intuitive control.

Mechanically, the robot will have to cope with force-depended friction in the Bowden cables. This issue will be addressed by a model-based controller with individual parameters for each axis.

Acknowledgment

The work is funded by the German Federal Ministry of Education and Research under grants 16SV5773 and 16SV5774.

References

[1] Y. Kwoh, *et al.*, "A new computerized tomography-aided robotic stereotaxis system," Robotics Age, 1985, pp. 17-22.

[2] J. Doll, W. Schlegel, O. Pastyr, V. Sturm, and W. Maier-Borst, "The use of an industrial robot as a stereotactic guidance system.," in CAR 87, Berlin, 1987.

[3] R. H. Taylor, *et al.*, "An Image-Based Robotic System for Hip Replacement Surgery," Journal of the Robotics Society of Japan, 1990, pp. 111-116.

[4] C. Caetano da Rosa, Operationsroboter in Aktion. Bielefeld, Germany: transcript Verlag, 2013.

[5] M. Jakopec, *et al.*, "The First Clinical Application of a "Hands-on" Robotic Knee Surgery System," Computer Aided Surgery, vol. 6, 2001, pp. 329-339.

[6] H. Mayer, I. Nagy, A. Knoll, E. U. Schirmbeck, and R. Bauernschmitt, "A robotic system providing force feedback and automation for minimally invasive heart surgery," in CARS, Osaka, Japan, 2006, pp. 265-267.

[7] M. Shoham, M. Burman, E. Zehavi, and Y. Kunicher, "MARS: miniature bone-mounted robot," in ISRACAS'2003, Tel Aviv, Israel, 2003.

[8] P. P. Pott, *et al.*, "ITD - A handheld manipulator for medical applications - Concept and design," in 3rd annual meeting of CAOS, Marbella / Spain, 2003, pp. 290-291.

[9] P. P. Pott, A. Wagner, E. Badreddin, P. Weiser, and M. L. R. Schwarz, "Inverse Dynamic Model and a Control Application of a Novel 6-DOF Hybrid Kinematics Manipulator," Journal of Intelligent and Robotic Systems, vol. 63, 2010, pp. 3-23.

[10] A. El-Shenawy, A. Wagner, P. P. Pott, and E. Badreddin, "Disturbance Attenuation of a Handheld Parallel Robot," in IEEE ICRA, Karlsruhe, 2013.

[11] T. Würtz, *et al.*, "The Twisted String Actuation System: Modeling and Control," in IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Montréal, Canada, 2010, pp. 1215-1220.

[12] K. Kong , J. Bae, and M. Tomizuka, "Control of Rotary Series Elastic Actuator for Ideal Force-Mode Actuation in Human-Robot Interaction Applications," IEEE/ASME Transactions on Mechatronics, vol. 14, 2009, pp. 105-118.

[13] M. Grün, R. Müller, and U. Konigorski, "Model Based Control of Series Elastic Actuators," in IEEE Biomedical Robotics and Biomechatronics (BioRob), Rome, Italy, 2012, pp. 538-543.

[14] N. Sadaghzadeh, L. Zouaghi, A. Wagner, J. Poshtan, and E. Badreddin, "Cascaded Error Estimation and Compensation of an Inertial Measurement Unit using the Fusion of Optical and Inertial Sensors," in COMADEM, Helsinki, 2013.

[15] E. Badreddin, "Recursive Control Structure for Mobile Robots," in International Conf. on Intelligent Autonomous Systems 2 (IAS.2), Amsterdam, 1989, pp. 11-14.

[16] A. Wagner, E. Nordheimer, and E. Badreddin, "Hierarchical constraint-based singularity avoidance," in System Theory, Control and Computing (ICSTCC), Sinaia, 2012.

[17] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," The International Journal of Robotics Research, vol. 5, 1986, pp. 90-98.

A Real-Time Architecture for Embodied Conversational Agents: Beyond Turn-Taking

Bahador Nooraei, Charles Rich and Candace L. Sidner
Computer Science Department
Worcester Polytechnic Institute
Worcester, USA
Email: bahador.nooraei@gmail.com, rich@wpi.edu, sidner@wpi.edu

***Abstract*—We describe the design, implementation and use of a middleware system, called DiscoRT, to support the development of virtual and robotic conversational agents. The use cases for this system include handling conversational and event-based interruptions, and supporting engagement maintenance behaviors, such as turn-taking, backchanneling, directed gaze and face tracking. The multi-threaded architecture of the system includes both "hard" and "soft" real-time scheduling and integrates with an existing collaborative discourse manager, called Disco. We have used the system to build a substantial conversational agent that is about to undergo long-term field studies.**

***Keywords*-engagement; interruption; turn-taking; backchannel; barge-in; schema; BML; arbitration.**

I. Introduction

Embodied conversational agents, both virtual and robotic, are becoming increasingly common and sophisticated in their interaction with humans. Originally, the transition from existing conversational agents (which typically interacted with users via alterating text input and output) to the "embodied" versions focused on animating a virtual face and/or body with expressions and gestures to complement the agent's utterances. More recently, however, conversational agents have acquired sensors, such as microphones, webcams and Kinects™, which are used to recognize users' speech and track their faces and bodies. Furthermore, many conversational agents now have *real* bodies, in the form of robots [1].

Along with these developments have come increasing expectations for more "natural" real-time interactions between agents and humans than the simple turn-taking of the earliest conversational agents. Unfortunately, the computational architectures underlying conversational agents have not evolved to keep up with the addition of sensors and bodies. Most such agent architectures, including several that we have built, have basically added an ad-hoc collection of real-time mechanisms to what is underlyingly the "letter" model of interaction discussed below. The goal of this work has been to make a fresh start and design a principled architecture that supports the continuous mutual signaling model discussed below.

Figure 1(a) shows the "letter" model of interaction, which still underlies most conversational agent implementations. In this model, one participant constructs an utterance (letter) containing the information to be communicated, sends it to the other participant, and waits. When the utterance is received, the other participant interprets it, thinks about the content, reacts

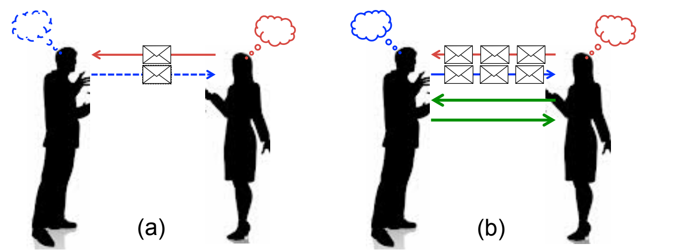


Figure 1. (a) Letter versus (b) continuous mutual signaling model.

(perhaps by performing an action) and then constructs a reply, if any, sends it back and waits. The cycle then repeats.

This is a very poor model of how co-present humans naturally interact. In reality, human conversation is a kind of "continuous mutual signaling," as shown in Figure 1(b). In this model, both participants are continuously and simultaneously producing, observing and responding to behaviors, both verbal and nonverbal. Furthermore, the exact timing (synchronization) of these behaviors is a crucial part of the interaction. Herbert Clark's keynote address at the AAAI 2010 Fall Symposium on Dialogue with Robots, titled "Talk and Its Timing," [2] explained this well and strongly motivated our work.

Thus communication in real conversations is multi-channel (multi-modal). The phenomenon of verbal turn-taking is just a strategy for managing the signal interference characteristics of one particular communication channel, namely audio (as compared to vision, for example, in which multiple signals can more easily co-exist). And in fact, people talk at the same time quite often, e.g., for barge-in (interrupting someone while they are talking) or verbal backchanneling (e.g, saying "Uh-huh" while someone else is talking).

We have designed, implemented and tested a new multi-threaded real-time architecture, called DiscoRT (Disco for Real-Time) to support the continuous mutual signaling model for embodied conversational agents. Disco is the name of the collaborative discourse (dialogue) manager that is a key component of the architecture—see Section V.

After first discussing related work in the next section, we will briefly introduce the Always-On project [3] which motivated the development of DiscoRT. In this project, we are using DiscoRT to develop a conversational agent, with both virtual and robotic embodiments, to provide social support for isolated older adults. Next, we will focus on nine key use cases

 ISBN: 978-1-61208-325-4

that guided our design. Following that, we will describe the implemented architecture and explain how it supports each of the use cases. Finally, in the conclusion, we will share our experience thus far in using the architecture.

II. Related Work

Broadly speaking, this work falls into the technical area of "behavior coordination mechanisms"—see review articles by Pirjanian [4] and by Scheutz and Andronache [5]. In particular, we are using what Pirjanian calls a priority-based arbitration model. The seminal example of such a model is Brooks' subsumption architecture [6]. Such models were a response to the observation that agents often have multiple and sometimes conflicting goals, i.e., goals that require different and incompatible behaviors.

We also use Arkin's [7] concept of parallel perception and action *schemas*, which is motivated in part by brain theory and psychology.

The Petri-net synchronization mechanism described in Section V-E is modeled on the Behavior Markup Language (BML) [8] developed for scripting the animation of the first generation of embodied virtual agents.

Other researchers whose work is most similar to ours include Thorisson [9], Bohus and Horvitz [10], and Chao and Thomaz [11]. Thorisson's architecture is more subsumption-like than ours, with three layers of "feedback loops" (schemas) in a fixed priority, whereas our system has only two layers and arbitrates between schemas with variable priorities. Bohus and Horvitz incorporate Bayesian statistics into their model, which we do not. Chao and Thomaz use Petri nets, as we do, to implement synchronization between multimodal behaviors, but do not support arbitration between conflicting behaviors. None of these systems includes a dialogue manager similar to Disco.

Finally, an important real-time issue that our architecture does not directly address is incrementality, especially with respect to natural language processing. For example, as soon as someone says, "the red ball...", before they finish the rest of the sentence, hearers will start to direct their gaze toward a red ball if it is easily visible in the immediate environment. Scheutz [12] and Traum et al. [13] have implemented this kind of incremental processing for interaction with robots and virtual agents.

III. The Always-On Project

Before presenting the use cases for DiscoRT below, we first briefly describe the Always-On project to provide a source of some concrete behavioral examples.

The Always-On project is a four-year effort, currently in its fourth year, supported by the U.S. National Science Foundation at Worcester Polytechnic Institute and Northeastern University. The goal of the project is to create a relational agent that will provide social support to reduce the isolation of healthy, but isolated older adults. The agent is "always on," which is to say that it is continuously available and aware (using a camera and infrared motion sensor) when the user is in its presence and can initiate interaction with the user, rather than, for example requiring the user to log in to begin interaction. Our goal is for

Figure 2. Embodied conversational agent in Always-On project.

the agent to be a natural, human-like presence that "resides" in the user's dwelling for an extended period of time. Beginning in the winter of 2014, we will be placing our agents with about a dozen users for a month-long, four-arm, evaluation study.

We are experimenting with two forms of agent embodiment. Our main study will employ the virtual agent Karen, shown in Figure 2, that comes from the work of Bickmore et al. [14] Karen is a human-like agent animated from a cartoon-shaded 3D model. She is shown in Figure 2 playing a social game of cards with the user. Notice that user input is via a touch-screen menu. Also, the speech bubble does not appear in the actual interface, which uses text-to-speech generation. We are also planning an exploratory study substituting the Reeti [15] tabletop robot, shown in Figure 3, for Karen but otherwise keeping the rest of the system as much the same as possible.

Figure 3. Reeti.

In total, the conversational agent or robot can interact with the user in more than ten different activities including: discuss the weather, learn about the agent, play a social game of cards, talk about family/friends, record a life story to the agent, promote exercise, promote nutrition, hear a humorous tale from the agent, get health tips from the agent, speak with a friend/family member via Skype™ (with all the details of Skype managed by the agent), and manage a personal calendar for the user.

A typical interaction with the agent might start with some greetings (specific to the time of day) and then some discussion of the weather. The weather discussion can be as short as today's weather forecast or extend to the next day, weather in other cities, and weather where friends or family live. At the user's choice, weather might be followed by a social game of cards where the agent's and user's hands in the game and the way the game is played out are commented upon. If the user and agent are somewhat well acquainted, thereafter might follow discussion of the user's family and friends.

IV. The Use Cases

The evaluation of software tools or middleware can be problematic for experimentally-focused disciplines, such as human-computer interaction. Straightforward application of the controlled experiment methodology suggests having two parallel development efforts implementing the same system requirements (one using the new software and one using something else) and then comparing measures of both the development process, such as time spent, and the performance of the resulting systems. Unfortunately, this approach is seldom practical, especially if the tool is designed for professional developers working on large, challenging systems. An experiment with a small, simple system is not likely to show the benefits of the tool.

We have therefore adopted a methodology from software engineering called *use cases* [16], which is particularly well-suited to the domain of interactive systems. In this approach, one identifies at the beginning of the design process a collection of archetypal behaviors (the “use cases”) that the system should support and then evaluates the implemented system in terms of how well it supports each of these behaviors. Furthermore, the use cases should be chosen to cover the most challenging aspects of the system’s required behavior.

We present nine such use cases for DiscoRT below and then return in Section VI to evaluate how our middleware architecture supports each of them. The point here is not that these use cases are unique to DiscoRT (these are well known behaviors) but rather that DiscoRT is the first agent architecture to support them in a principled, unified design.

Notice that the design of DiscoRT aims to support virtual agents and robots that, unlike in the Always-On project, can use their hands and arms to point at and manipulate objects in their environment. DiscoRT is also designed to support fully spoken interaction, as well as the menu-based interaction mode of the Always-On project.

A. Engagement Use Cases

The first group of six use cases for DiscoRT concern supporting *engagement* between the agent and the user. As defined by Sidner et al. [17], engagement is “the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake.” Furthermore, as we will see below, engagement involves both verbal and (a lot of) nonverbal behavior.

Case 1) Walking Up to the Agent: The first use case relates to establishing engagement. Specifically, when the user walks by and triggers the motion detector, the agent should awake out of its quiescent state and issue a greeting, such as “good morning.” If the user then responds by approaching the computer (which the agent can notice with face detection software on the webcam), the agent should continue start face tracking (see next use case) and continue with a further engaging utterance, such as “how did you sleep?”

The next four use cases relate to maintaining engagement. In later work, we identified four types of what they called “connection events” that function to maintain engagement [18], [19]. We have a use case corresponding to each type of connection event. In general, these use cases involve crucial timing

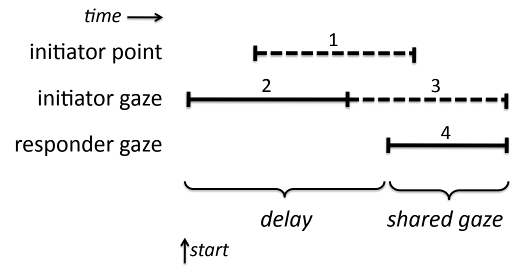


Figure 4. Time line for directed gaze.

constraints between the verbal and/or nonverbal behaviors of the user and the agent.

Case 2) Face Tracking: Face tracking is the agent’s attempt to achieve what is technically called *mutual facial gaze* [20] with the user. When the agent is face tracking, it should orient its gaze toward where it detects the user’s face. In addition to being the agent’s default gaze behavior for maintaining engagement, mutual facial gaze can have other interaction functions. For example, it is typical to establish mutual facial gaze at the end of a speaking turn (see next use case).

Case 3) Turn-Taking: Even though, as discussed in Section I, a conversational interaction entails much more than turn-taking, an embodied conversational agent nevertheless does need to manage speaking turns, particularly in a menu-based system. In linguistics, an *adjacency pair* [21] is the term used to refer to two utterances by two speakers, with minimal overlap or gap between them, such that the first utterance provokes the second utterance. A question-answer pair is a classic example of an adjacency pair. Thus, after producing the so-called “first turn” of an adjacency pair, the agent should wait until the user responds (or until some specified timeout occurs). In some conversational circumstances, the user’s response can also be followed by a “third turn” in which the agent, for example, acknowledges the user’s response.

Importantly, we generalize the concept of adjacency pair beyond the traditional linguistic definition to include both verbal and nonverbal responses. So for example, a nod can be the answer to a question, instead of a spoken “yes,” or the performance of an action can be the nonverbal response to a verbal request, such as, “please pass the salt.” Adjacency pairs, of course, also often overlap with the other nonverbal behaviors, such as face tracking and directed gaze (see Use Case 5).

Case 4) Backchanneling: A *backchannel* is an interaction event in which a listener directs a brief verbal or nonverbal communication back to the speaker *during* the speaker’s utterance. Typical examples of backchannels are nods and/or saying “uh, huh.” Backchannels are typically used to communicate the listener’s comprehension of the speaker’s communication (or lack thereof, e.g., a quizzical facial expression) and/or desire for the speaker to continue. A conversational agent should be able to both generate appropriate backchannels and interpret backchannels from the user. For example, in the Always-On life-story recording activity, the agent should nod appropriately, even though it does not understand the content of the story that being recorded.

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-325-4

Case 5) Directed Gaze: Finally, Figure 4 shows the time line for the last, and most complex, engagement maintenance use case, called *directed gaze* [22]. In this behavior, one person (the *initiator*) looks and optionally points at some object in the immediate environment in order to make it more salient, following which the other person (the *responder*) looks at the same object. This behavior is often synchronized with the initiator referring to the object(s) verbally, as in "now spread the *cream cheese* on the *cracker*" (pointing first to the cream cheese and then to the cracker). By turning his gaze where directed, the responder intends to be cooperative and thereby signals his desire to continue the interaction (maintain engagement).

In more detail (see Figure 4), notice first that the act of pointing (1), if it is present, begins after the initiator starts to look (2) at the object. (This is likely because it is hard to accurately point at something without looking to see where it is located.) After some delay, the responder looks at the salient object (4). The initiator usually maintains the pointing (1), if it is present, at least until the responder starts looking at the object. However, the initiator may stop looking at the object (2) before the responder starts looking (4), especially when there is pointing. (This is often because the initiator looks at the responder's face, assumedly to check whether the responder has directed his gaze yet.) Finally, there may be a period of shared gaze, i.e., a period when both the initiator (3) and responder (4) are looking at the same object.

Case 6) Walking Away: The last use case in this section relates to ending engagement. Hopefully, most of the time disengagement between the agent and user will occur as the result of an explicit leave-taking conversational exchange, such as, "Goodbye; See you later." However, the agent should also be prepared to deal with the user simply walking away at any time.

B. Interruption Use Cases

The remaining three use cases relate to various kinds of interruption behaviors. The ability to do more than one thing at a time and smoothly shift between them is a key requirement for a natural conversational agent.

Case 7) Scheduled Event: One reason for interrupting an activity is to due the (imminent) occurrence of a scheduled event. For example, in the Always-On project, the agent helps the user keep a personal calendar of events such as lunch dates, doctor appointments, etc. If the user has a lunch date at noon, the agent should interrupt whatever the agent and user are doing together (e.g., playing cards) ten or fifteen minutes before noon to remind the user of the lunch date and to wrap up or postpone the current activity.

Case 8) Changing Topic: A conversational agent should be able to, either of its own volition, or in response to the user's behavior, smoothly change the topic of the conversation and then, if appropriate, smoothly return to the original interrupted topic. For example, in the Always-On project, activities such as playing cards are viewed as social "containers," within which other topics can also be discussed. In one of our target scenarios, at the end of the user's turn in a card game, the agent says, "By the way, have you thought about my suggestions for how to get more exercise?" After a short discussion of exercise, the agent returns to the card game by saying, "Ok, so I think it's your turn now."

Case 9) Barge-In: Barge-in is a common conversational phenomenon similar to backchanneling (Case 4), in that the listener starts communicating before the speaker's turn is finished. In the case of barge-in, however, the listener's intention is for the speaker to stop and let the listener "take" the turn. A conversational agent should be able to respond to the user's barge-in by promptly ceasing to talk. In a purely spoken language system, the user can barge in simply by starting to speak. A menu-driven system, such as the Always-On project, can support user barge-in by displaying a menu for the user to click on while the agent is still speaking.

The case of a conversational agent barging in on the user is less common. However, Chao and Thomaz [23] describe a strategy, called *minimum necessary information*, for a robot to start acting before the user has finished speaking instructions.

V. System Architecture

Figure 5 shows the architecture of the DiscoRT system that we have designed, implemented and used in the Always-On project to support the use cases above in a principled and general way. Four key features of this architecture are:

1) Multiple Threads: Supporting the continuous mutual signaling model of interaction obviously requires a highly parallel architecture with multiple threads. Each input and output modality and the internal decision making processes needs to function independently without blocking one other.

2) Resource Arbitration: We think of the agent's face, voice, hands and gaze as resources that can be used for different (and sometimes competing) purposes in an interaction. For example, the agent's gaze can be used to achieve mutual facial gaze (Case 2) or it can be used direct the user's gaze to a salient object in the environment (Case 5). Similar to a computer operating system, one of the key functions of DiscoRT is to arbitrate between competing demands for a given resource.

3) Real-Time Control: Timing is critical in all of the use cases. However, there is more margin for error in some use cases than in others. For example, an inappropriate delay of even a fraction of a second in the agent's response to a user-initiated directed gaze or barge-in will be noticeable and degrade the believability of the agent. On the other hand, changing topics or reminding the user of an upcoming scheduled event can be delayed a second or two without harmful effect. We call these "hard" and "soft" real-time constraints, respectively, and handle them separately in the architecture.

4) Dialogue Management: A unique feature of our architecture is its integration with the Disco dialogue manager. Among other things, Disco provides a focus stack for helping to manage interruptions.

We return now to Figure 5 for a high-level tour of the architecture, following which we will discuss several aspects in more detail.

Starting at the left of the figure, we see the *perceptors*, which implement the input side of the system's multimodal interface with the user. At the bottom of the figure are the resources which the system needs to control/manage, such as

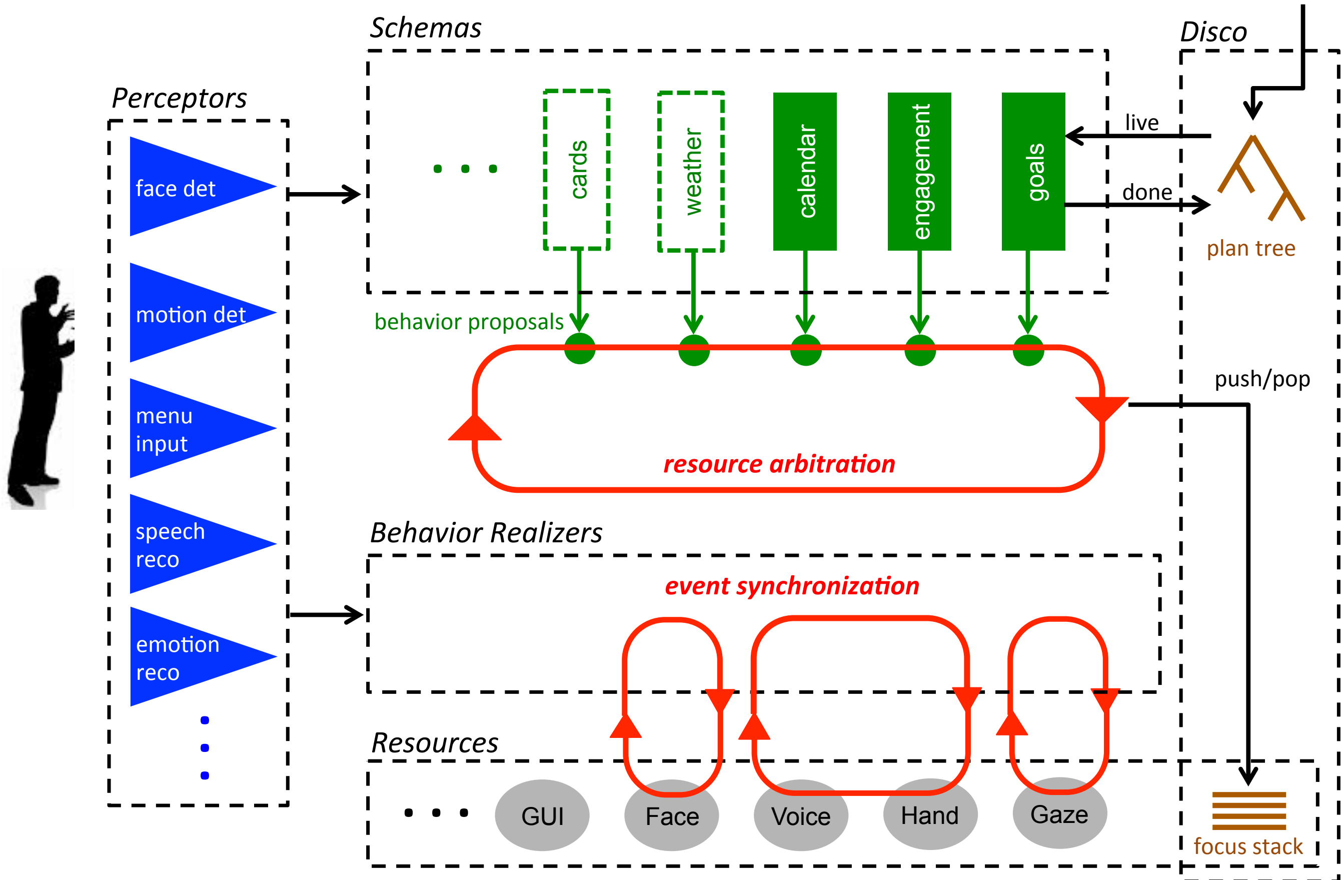


Figure 5. DiscoRT system architecture.

the agent's face, voice, hands, gaze and so on (the specific resources may vary between agents). Notice that Disco's focus stack is also viewed as a resource; the implications of this will be discussed below. Along the top of the figure are an open-ended collection of *schemas*, which in parallel make proposals for behaviors using particular resources. Resource conflicts between behavior proposals are arbitrated by the *resource arbitration* (soft) real-time process—the loop in the middle of the figure. Each behavior that survives the arbitration process results in instantiation of a *realizer* process(the loops in the figure that intersect the resources) that handles the (hard) real-time synchronization of behavioral events involving the assigned resource(s). Finally, the right side of the figure shows the Disco dialogue manager, which has a plan tree and focus stack as its two main data structures.

A. *Perceptors*

Perceptors are the system's abstraction for sensing capabilities, such as face detection, motion detection, menu input, speech recognition, and so on. The specific perceptors may vary between agents, depend on the hardware and software configuration of the agent.

Some perceptors simply abstract the output of a single hardware sensor, such as the infrared motion detector in the Always-On project. However, perceptors can also fuse information from multiple input modes, such as an emotion recognition perceptor that combines facial expression information from a camera with tone of voice information from a microphone. A single item of hardware, such as a camera, may also serve multiple perceptors, such as face detection and gaze tracking. In the Always-On project, the touch-screen menu is also modeled as a perceptor.

Each perceptor runs on its own thread with a loop cycle rate depending on its needs, such as the frame rate of the camera. The output of the perceptors is used by both the schemas and the realizers (discussed in more detail below). For example, face detection and motion perception are used by the schema that establishes engagement. Gaze tracking is used by the realizer that implements the direct-gaze behavior discussed in Case 5. Perceptors support both polling and event-based API's.

B. *Schemas*

Schemas are the core of the DiscoRT architecture. In the Always-On project there is a schema corresponding to each social activity that the agent can do with the user, such as talking about the weather, playing cards, etc. These schemas are created and destroyed as the system runs. A few other schemas, such as the schema that manages engagement, are always running.

The fundamental function of a schema is to continually propose *behaviors*. Each schema runs on its own thread with a

 ISBN: 978-1-61208-325-4

loop cycle rate depending on its needs and typically maintains its own internal state variables. In order to decide what behavior to propose at a given moment (it is allowed to propose no behavior), a schema can consult its own state variables and the system's perceptors, as well as external sources of information. For example, the weather schema downloads up-to-date weather information from the internet.

Conceptually, a behavior specifies a "program" (called the *realizer*—see Section V-E) to be executed using one or more resources. For example, when a schema wants to "say" something, it proposes a text-to-speech behavior that specifies the utterance string and requires the voice resource. A more complex example is the directed gaze behavior (see Figure 4), which requires three resources (voice, hand and gaze) and includes a complex realizer program to synchronize the control of the resources with input from the perceptors.

Some schemas, such as for talking about the weather, are purely conversational, (see Section V-D), and some, such as playing cards, also involve the manipulation of shared artifacts, such as cards. In the case of a virtual agent, such manipulations use the *GUI* resource. In the Always-On project, playing cards is a social activity, that involves both manipulation and conversation about the game, e.g., "I've got a terrible hand."

There are no restrictions in DiscoRT on the internal implementation of a schema. In the Always-On project we have primarily implemented schemas as state machines and using D4g [24], which is an enriched formalism for authoring dialogue trees.

C. Resource Arbitration

The resource arbitration thread/loop runs approximately once per second and gathers up the collection of behavior proposals from all of the running schemas. Proposals with non-overlapping resource requirements are directly scheduled, i.e., an instance of the specified realizer is created and started running (unless it is already running). Resource conflicts between proposals are resolved using a simple system of per-schema priorities with some fuzzy logic rules [25] to make the system more stable, i.e., to prevent switching behaviors too quickly between closely competing schemas. The behavior proposals that are chosen are then scheduled.

D. Dialogue Management

Since DiscoRT is designed to support conversational agents, it includes specialized machinery for dialogue management. We are using the Disco dialogue manager, which is the open-source successor to Collagen [26], [27]. Disco has two key data structures: a plan tree and a focus stack. Each of these has a point of integration with DiscoRT, as shown in Figure 5.

The *plan tree*, which is typically provided from another component outside of DiscoRT, represents the agent's goals for its interaction with user as a hierarchical task network [28]. This formalism includes optional and repeated goals and partial ordering constraints between subgoals. Thus the typical interaction example, starting with greetings, etc., described in Section III above is formalized as a plan tree. The plan tree can be updated while the system is running.

DiscoRT includes a predefined schema, called the *goals* schema in Figure 5, that is always running and automatically starts the schema(s) corresponding to the currently live goal(s) in the plan tree. Thus, for example, when the "discuss the weather" goal becomes live, the goals schema starts the weather schema. When the schema exits, the corresponding goal in the plan tree is automatically marked as done.

The *focus stack* is stack of goals that captures the familiar phenomenon of pushing and popping topics in human conversation. (Technically, it is stack of focus spaces, each of which includes a goal, but this is beyond the scope of this paper.) For example, in the middle of a card game, the agent might temporarily suspend (push) playing cards to talk about exercise or to remind the user about an upcoming appointment in her calendar, and then return (pop) to playing cards.

In DiscoRT, this kind of interaction is achieved by making the focus stack a resource that represents control of the current topic of conversation. Schemas that involve conversation, such as the cards schema, the exercise schema, and the calendar schema in the example above, require the focus stack in their behavior proposals. When the resource arbitrator starts a new behavior realizer that requires the focus stack, it pushes the goal associated with the proposing schema onto the stack (unless it is already there). When the behavior realizer finishes, the goal is automatically popped off the stack. (See Case 7 in Section VI for detailed example.)

This integration between the dialogue model and the schema architecture in DiscoRT is very powerful and flexible. For example, it is possible and sometimes useful for a schema to propose a speech behavior *without* requiring the focus stack. If you poke a robot, it might respond by saying "Ouch!" without changing the conversational focus. We are also experimenting with providing hooks in DiscoRT for automatically producing generic transition language when the stack is pushed and popped, such as "excuse the interruption, but..." or "now, returning to..."

E. Behavior Realizers

The behavior realizers (bottom of Figure 5) implement the hard real-time event synchronization in the system, such as responding to user-initated directed gaze. A behavior realizer in DiscoRT is very similar (hence the name) to a BML realizer. However, for the reasons discussed in detail in [29], we use an event-driven Petri net rather than a fixed schedule as in most BML realizers.

Further, in DiscoRT there can be multiple realizers independently controlling non-overlapping resource sets (think of a robot "rubbing its tummy and patting its head" at the same time). Each behavior realizer has a separate thread that runs by default at 10Hz. Realizers often get information from perceptors.

A simple realizer, such as for a smile behavior, uses only one resource (e.g., face) and does not get information from perceptors (e.g., it just waits until the smile is completed). More complex realizers, such as for a directed gaze behavior, use multiple resources (e.g., voice, hand and gaze) and access one or more perceptors (e.g., hand and gaze tracking) to implement multimodal synchronization.

A realizer starts execution when the resource arbitrator accepts a schema's behavior proposal as described above. If the schema stops proposing that behavior, then the realizer is automatically stopped. The realizer program can also stop itself (e.g., if has completed the behavior), in which case the proposing schema is notified, so that it can update its internal state for making future proposals.

Our realizers support all of the essential timing relationships of BML (synchronize, before, after and fixed delay). We did not implement the complete BML specification because much of it concerned specific gestural actions; we wanted a more general framework. Also, our realizer programming API uses Java rather than XML.

VI. Evaluation of the Use Cases

In this section, we describe how each of the use cases defined in Section IV is handled by the DiscoRT architecture.

Case 1) Walking Up to the Agent: The engagement schema (implemented as a state machine) continually polls the motion perceptor (which abstract the IR motion detector hardware). When motion is detected, the schema proposes a speech behavior (e.g., "Good morning") and enters a state in which it starts polling the face perceptor (which abstracts the output of face detection software operating on the webcam output). When a face is detected, the schema proposes another speech behavior and enters the *engaged* state; otherwise, after a timeout, the schema returns to the motion perceptor polling state.

Case 2) Face Tracking: Face tracking is implemented by a behavior realizer that requires the gaze resource and uses the face detection perceptor. The realizer simply runs a loop that updates the agent's gaze to where it currently sees the user's face. The face tracking behavior that causes the realizer to be started is proposed by the engagement schema when it enters the engaged state (see Case 6 for stopping the realizer).

Case 3) Turn-Taking: Turn-taking is implemented by an abstract state machine that is reused in all of the schemas that include conversational behavior, such as weather, cards and calendar. The agent's utterances are produced by proposing speech behaviors. In menu-based systems, the state machine waits for the user's response by waiting for an event from the menu perceptor (which abstracts the menu GUI) or the speech perceptor (which abstracts speech recognition).

Case 4) Backchanneling: Backchanneling is a hard real-time phenomenon, so it must be implemented by a behavior realizer that receives events from a perceptor that detects appropriate moments at which to produce backchannels, such as the end of phrases or sentences in the user's speech. The schema that proposes the backchanneling behavior is responsible for deciding what form of backchannel should be used (e.g., positive or negative) and for re-proposing a new behavior when the form of backchannel should be changed.

Case 5) Directed Gaze: As discussed in Section V-E, the time line for directed gaze (Figure 4) is implemented as a behavior realizer. Any schema can propose a directed gaze behavior.

Case 6) Walking Away: The engagement schema state machine includes a timeout to notice when there has been no face or motion perceived. When this occurs, it stops proposing face tracking and enters the waiting for engagement state described in Case 1.

Case 7) Scheduled Event: While the user and agent are playing cards together, both the card schema and the calendar schema are running, but the calendar schema is not making any behavior proposals. The card schema's behavior proposals require the focus stack, so the card playing goal stays on the top of the stack. Then, triggered by the clock time, the calendar schema proposes a speech behavior that requires the focus stack. Because the calendar schema has a higher priority, the arbitrator gives it control of the focus stack, which causes the calendar reminder goal to be pushed on top of the card playing goal. The calendar reminder goal remains on the top of the stack throughout the (sub)dialogue regarding the upcoming appointment. When this reminder dialogue is completed, the calendar schema stops making proposals, the calendar goal is popped, and the arbitrator gives the focus stack resource back to the card schema, which has been continuously proposing the next behavior in the game, but never getting the needed focus stack resource.

Case 8) Changing Topic: Changing topic is handled similarly to Case 7, except that instead of the calendar schema deciding to make the interruption, the decision that the new topic has a higher priority than the current topic is made by some other schema based on cognitive reasoning that is outside of the scope of DiscoRT. The mechanism of pushing and popping the focus stack is identical, however. Furthermore, the interrupted schema may stop proposing the old topic, in which case it is never returned to.

Case 9) Barge-In: Barge-in is handled in the speech behavior realizer, which in addition to controlling the text-to-speech engine (voice resource), listens for events from the menu or speech perceptor (depending on the type of system). When such an event is received, the realizer immediately stops the text-to-speech engine and terminates.

VII. Conclusion

DiscoRT has succeeded in its design goal of supporting the specified use cases in a principled and general way and has been extremely useful in implementing the Always-On project. Furthermore, we feel that the use cases themselves are a research contribution towards evaluating other systems for similar purposes. Our only negative report on DiscoRT is that it has been challenging to master the schema-based programming approach. This is not entirely surprising, since programming highly-parallel systems is well-known to be difficult. DiscoRT is implemented in Java and is available from the authors under an open-source license.

Acknowledgment

This work is supported in part by the National Science Foundation under award IIS-1012083. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] C. Rich and C. Sidner, "Robots and avatars as hosts, advisors, companions and jesters," AI Magazine, vol. 30, no. 1, 2009, pp. 29–42.

[2] [Online]. Available: http://hci.cs.wisc.edu/aaai10

[3] [Online]. Available: http://www.cs.wpi.edu/~rich/always

[4] P. Pirjanian, "Behavior coordination mechanisms: State-of-the-art," Inst. for Robotics and Intelligent Systems, U. of Southern California, Tech. Rep. IRIS-99-375, 1999.

[5] M. Scheutz and V. Andronache, "Architectural mechanisms for dynamic changes of behavior selection strategies in behavior-based systems," IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 34, no. 6, 2004, pp. 2377–2395.

[6] R. Brooks, "A robust layered control system for a mobile robot," IEEE J. of Robotics and Automation, vol. 2, 1986, pp. 14–23.

[7] R. Arkin and D. MacKenzie., "Temporal coordination of perceptual algorithms for mobile robot navigation," Robotics and Automation, vol. 10, no. 3, 1994, pp. 276–286.

[8] S. Kopp et al., "Towards a common framework for multimodal generation: The Behavior Markup Language," in Proc. Int. Conf. on Intelligent Virtual Agents, 2006, pp. 205–217.

[9] K. Thorisson, "Natural turn-taking needs no manual: Computational theory and model, from perception to action," Multimodality in Language and Speech Systems, vol. 19, 2002, pp. 173–207.

[10] D. Bohus and E. Horvitz, "Models for multiparty engagement in open-world dialog," in Proceedings of the SIGDIAL 2009 Conference. London, UK: Association for Computational Linguistics, September 2009, pp. 225–234.

[11] C. Chao and A. Thomaz, "Timing in multimodal turn-taking interactions: Control and analysis using timed petri nets," J. Human-Robot Interaction, vol. 1, no. 1, 2012, pp. 4–25.

[12] E.Krause, R. Cantrell, E. Potapova, M. Zillich, and M. Scheutz, "Incrementally biasing visual search using natural language input," in Proc. 12th Int. J. Conf. on Autonomous Agents and Multiagent Systems, St. Paul, MN, 2013, pp. 31–38.

[13] D. Traum, D. DeVault, J. Lee, Z. Wang, and S. Marsella, "Incremental dialogue understanding and feedback for multiparty, multimodal conversation," in Proc. Int. Conf. on Intelligent Virtual Agents, Santa Cruz, CA, 2012, pp. 245–251.

[14] T. Bickmore, L. Caruso, and K. Clough-Gorr, "Acceptance and usability of a relational agent interface by urban older adults," in Proc. ACM Conf. on Computer Human Interaction. ACM, 2005, pp. 1212–1215.

[15] [Online]. Available: http://www.reeti.fr

[16] D. Leffingwell and D. Widrig, Managing Software Requirements: A Use Case Approach (Second Edition). Addison-Wesley, 2012.

[17] C. L. Sidner, C. Lee, C. Kidd, N. Lesh, and C. Rich, "Explorations in engagement for humans and robots," Artificial Intelligence, vol. 166, no. 1-2, 2005, pp. 104–164.

[18] C. Rich, B. Ponsler, A. Holroyd, and C. Sidner, "Recognizing engagement in human-robot interaction," in Proc. ACM Conf. on Human-Robot Interaction, Osaka, Japan, Mar. 2010, pp. 375–382.

[19] A. Holroyd, C. Rich, C. Sidner, and B. Ponsler, "Generating connection events for human-robot collaboration," in IEEE Int. Symp. on Robot and Human Interactive Communication, Atlanta, GA, 2011, pp. 241–246.

[20] M. Argyle and M. Cook, Gaze and Mutual Gaze. New York: Cambridge University Press, 1976.

[21] D. Crystal, The Cambridge Encyclopedia of Language. Cambridge, England: Cambridge University, 1997.

[22] A. Kendon, "Some functions of gaze direction in two person interaction," Acta Psychologica, vol. 26, 1967, pp. 22–63.

[23] C. Chao, C. Lee, M. Begum, and A. Thomaz, "Simon plays simon says: The timing of turn-taking in an imitation game," in Int. Symp. on Robot and Human Interactive Communication (RO-MAN), 2011, pp. 235–240.

[24] C. Rich and C. L. Sidner, "Using collaborative discourse theory to partially automate dialogue tree authoring," in Proc. Int. Conf. on Intelligent Virtual Agents, Santa Cruz, CA, Sep. 2012, pp. 327–340.

[25] J. Fodor and M. Roubens, Fuzzy Preference Modeling and Multicriteria Decision Support. Kluwer Academic, 1994.

[26] C. Rich and C. Sidner, "Collagen: A collaboration manager for software interface agents," User Modeling and User-Adapted Interaction, vol. 8, no. 3/4, 1998, pp. 315–350, reprinted in S. Haller, S. McRoy and A. Kobsa, editors, *Computational Models of Mixed-Initiative Interaction*, Kluwer Academic, Norwell, MA, 1999, pp. 149–184.

[27] C. Rich, C. Sidner, and N. Lesh, "Collagen: Applying collaborative discourse theory to human-computer interaction," AI Magazine, vol. 22, no. 4, 2001, pp. 15–25, Special Issue on Intelligent User Interfaces.

[28] C. Rich, "Building task-based user interfaces with ANSI/CEA-2018," IEEE Computer, vol. 42, no. 8, Aug. 2009, pp. 20–27.

[29] A. Holroyd and C. Rich, "Using the behavior markup language for human-robot interaction," in Proc. ACM Conf. on Human-Robot Interaction, Boston, MA, Mar. 2012, pp. 147–148.

Communicative Capabilities of Agents for the Collaboration in a Human-Agent Team

Mukesh Barange, Alexandre Kabil, Camille De Keukelaere, Pierre Chevaillier
ENIB (UEB), Lab-STICC
Brest, France
{barange, kabil, dekeukelaere, chevaillier}@enib.fr

Abstract—**The coordination is an essential ingredient for the human-agent teamwork. It requires team members to share knowledge to establish common grounding and mutual awareness among them. In this paper, we propose a behavioral architecture *C^2BDI* that allows to enhance the knowledge sharing using natural language communication between team members. We define collaborative conversation protocols that provide proactive behavior to agents for the coordination between team members. We have applied this architecture to a real scenario in a collaborative virtual environment for training. Our solution enables users to coordinate with other team members.**

Keywords-Human interaction with autonomous agents, Cooperation, Dialogue Management, Decision-Making

I. INTRODUCTION

In collaborative virtual environments (VE) for training, human users, namely learners, work together with autonomous agents to perform a collective activity. The educational objective is not only to learn the task, but also to acquire social skills in order to be efficient in the coordination of the activity with other team members [1]. Effective coordination improves productivity, and reduces individual and team errors. The ability to coordinate one's activity with others relies on two complementary processes: common grounding [2] and mutual awareness [3]. Common grounding leads team members to share a common point about their collective goals, plans and resources they can use to achieve them [2]. Mutual awareness means that team members act to get information about others' activities by direct perception, information seeking or through dialogues, and to provide information about theirs [3].

The collaboration in a human-agent teamwork poses many important challenges. First, there exists no global resource that human team members and virtual agents can rely on to share their knowledge, whereas, in a team of autonomous agents, the coordination can be achieved through the means of a mediator, or blackboard mechanism. Second, the structure of the coordination between human-agent team members is open by nature: virtual agents need to adopt the variability of human behavior, as users may not necessarily strictly follow the rules of coordination. In contrast, in agent-agent interactions, agents follow the rigid structure of coordination protocols (e.g., contract net protocol). Thus, the ability to coordinate with human team members requires to reason about their shared actions, and situations where team members need the coordination to progress towards the team goal. Moreover, another important characteristic of the human-human teamwork is that the team members pro-actively provide information needed by other team members based on the anticipation of other's needs of information [4]. Thus, in a human-agent team, agents should allow human team members to adjust their autonomy and help them to progress in their task.

The paper focuses on the task-oriented, collaborative conversational behavior of virtual agents in a mixed human-agent team. Other aspects of embodied virtual agents, such as emotions, facial expressions, non-verbal communication, etc. are out of the scope of this study. As the team members must have the shared understanding of skills, goals and intentions of other team members, we proposed a belief-desire-intention based (BDI-like) agent architecture named as *Collaborative-Conversational BDI agent architecture* (C^2BDI). On the one hand, this architecture provides the deliberative behavior for the realisation of collective activity and, on the other hand, it provides conversational behavior for the dialogue planning to exhibit human like natural language communication behavior for coordination. The contributions of this paper include: (1) the definition of collaborative communication protocols to establish mutual awareness and common grounding among team members; and (2) a decision-making mechanism where dialogues and beliefs about other agents are used to guide the action selection mechanism for agents to collaborate with their team members. The approach consists in formalizing the conversational behavior of the agent related to the coordination of the activity, which reduces the necessity to explicitly define communicative actions in the action plan of the agent. It also makes the human-agent interaction more adaptive.

In section II, we present related work on human-agent teamwork. Section III presents different components of our architecture. The conversational behavior is detailed in section IV. The next section illustrates how the solution fulfils the requirements of real educational scenarios. Finally, section VI summaries our positioning.

II. RELATED WORK

Both AI and dialogue literature agree upon the fact that to coordinate their activities, agents must have the joint-intention towards the group to achieve collective goal [5] and must agree upon the common plan of action [6]. The joint-intention theory specifies that agents have common intentions towards the group goal [5]. This theory does not guarantee that agents follow the same action plan. Comparing to this theory, the shared-plan theory [6] specifies that even agents share a common action plan to achieve the group goal, it does not guarantee that agents have the commitment towards the

 ISBN: 978-1-61208-325-4

group to achieve that goal. Both of these theories are mainly applied for the coordination among a group of artificial agents. The C^{2}BDI architecture takes the advantage of both of these theories to establish common grounding and mutual awareness among mixed human-agent team members.

A number of human-agent team models have been proposed in the literature [7], [8]. Collagen agent [7] is built upon the human discourse theory and can collaborate with a user to solve domain problems, such as planning a travel itinerary and user can communicate with agents by selecting the graphical menus. In [8], collaboration in teams is governed by teamwork notification policies, that is, when an important event occurs, the agent may notify the user with respect to appropriate modality and the user's position. To achieve collaboration between team members, [9] proposed a four stage model that includes (i) recognition of the potential for cooperation, (ii) team formation (iii) plan formation, and (iv) plan execution. Based on this model, [10] proposed an agent model and defines how collective intentions from the team formation stage are built up from persuasion and information-seeking speech act based dialogues, using motivational attributes goal and intention. Moreover, [11] proposed an agent based dialogue system by providing dialogue acts for collaborative problem solving between a user and a system. Recently, [12] have proposed a theoretical framework for proactive information exchange in agent teamwork to establish shared mental model using shared-plan [6].

One of the prominent approaches for dialogue modelling is the information state (IS) approach [13]. The IS defined in [14] contains contextual information of dialogue that includes dialogue, semantic, cognitive, perceptual, and social context. This model includes major aspects to control natural language dialogues. However, it does not include contextual information about the shared task. This leads to an incoherence between dialogue context and shared task in progress. In [15], an IS based interaction model for *Max* agent has been proposed that considers coordination as an implicit characteristic of team members. Comparing with [15], C^{2}BDI agents exhibit both reactive and proactive conversational behaviors, and explicitly handle cooperative situations through communication between team members. Moreover, [14] proposed a taxonomy of dialogue acts (DIT++) based on the dialogue interpretation theory. The semantics of these dialogue acts are based on the IS based approach. This taxonomy was built mainly to annotate natural language dialogues. We are motivated to use it to understand and interpret conversation between human-agent team due to its following characteristics: (i) it is mainly used for dialogue interpretation in human-human conversation; (ii) it supports task oriented conversation; and (iii) it has become the ISO 24617-2 international standard for dialogue interpretation using dialogue acts.

III. C^{2}BDI Agent Architecture

In this section, we describe components of C^{2}BDI agent architecture that provide deliberative and conversational behaviors for collaboration (see Fig. 1). We consider that C^{2}BDI agents are situated in an informed VE where agents can perceive entities and can access specific properties, such as the state, position, attribute values etc. of entities within their field

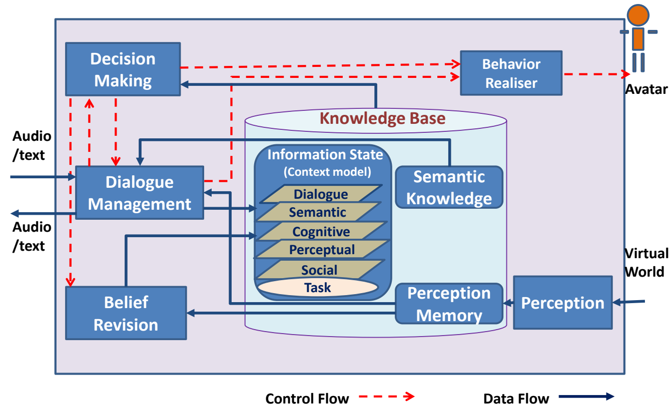


Figure 1: Components of Agent architecture and data flow

of perception. The agent architecture is based on the theory of shared-plan [6] and joint-intention [5].

The agent perceives the VE through the perception module. The current perceived state of the VE is an instantiation of concepts the agent holds in its semantic knowledge. The perception (in our case, multi-modal perception through vision and dialogue) allows agents to enrich their knowledge, and to monitor the progress of the shared activity. Agents have partial beliefs about the state of VE as they have limited perception. The belief revision specialises the classical belief revision function of BDI approach. Since, the state of the world can be changed due to an interaction by team members, the belief revision function periodically updates the knowledge base of the agent, and maintains the consistency of the knowledge base. The dialogue manager allows agents to share their knowledge with other team members using natural language communication. It supports both reactive and proactive conversation behavior, and ensures the coordination of the activity. The decision-making uses private beliefs and beliefs about others from the knowledge base to decide whether to elaborate the plan, identifying collaborative situations, to react to the current situation, or to exchange information with other team members. The behavior realiser module is responsible for the execution of actions and the turn taking behavior of the agent.

A. Knowledge Organisation

The organisation of knowledge in C^{2}BDI agent allows to establish the strong coupling between decision making and the collaborative conversational behavior of the agent. The knowledge base consists of semantic knowledge, perception memory and IS. The semantic knowledge contains semantic information that is known a priori by the agent, such as the knowledge concerning concepts, and individual and shared plans. Following the shared-plan theory [6], C^{2}BDI agents share the same semantic knowledge about the VE and the group activity. This simplifies the planning process of agents, as agents need to construct only their local plan. Moreover, sharing the same semantic knowledge also supports proactive conversation behavior of the agent as it allows the decision-making process to identify collaborative situations and information needed by other team members. The perception memory acquires information about the state of the VE perceived by the perception module, whereas, the IS contains contextual information about the current activity and dialogues.

 ISBN: 978-1-61208-325-4

B. Information State

The IS is primarily used in literature to control natural language dialogues [13], [14]. We extended its usage as the source of knowledge between the decision-making and conversational behavior of the C^2BDI agent to establish coherence between these two processes. In C^2BDI agent, the IS works as an *active memory* that contains beliefs and intentions of the agent. In C^2BDI agent, the semantic context of the IS is instantiated from concepts the agent holds in semantic knowledge, depending on the progress of the shared task. It includes the *agenda* that contains dialogue goals. These goals are added to the agenda due to communicative intentions generated by the realisation of the collaborative task and by the social obligations. To cooperate with other team members, the agent needs not only the information about the current context of the collective activity, but also beliefs about team members to establish common grounding and mutual awareness. To acquire these information, we extend the IS based context model of [14] by adding the *task context* to it (see Fig. 2).

Dialogue Context	agent-dialogue-acts, addressee-dialogue-acts, dialogue-act-history, next-moves	
Semantic Context	agenda, qud, communication-plan, beliefs, expected-dialogue-acts	
Cognitive Context	mutual-belief	
Social Context	communication-pressure	
Perception Context	object-in-focus, agent-in-focus, third-person-in-focus	
Task Context	cooperative-info	group-goal, group-desire, group-intention joint-goal, joint-desire, joint-intention, joint-commitment
	task	task-focus, goals, desires

Figure 2: Extended Information State in C^2BDI architecture

The *task context* of our IS includes information about the *task* that contains intentions *task-focus*, goals, and desires of the agent. The C^2BDI agent follows the theory of joint-intention [5] to ensure that each team member has a common intention towards the team goal, therefore, the *task context* also contains *cooperative-information*, which includes beliefs about *group-goal*, *group-desire*, *group-intention*, *joint-goal*, *joint-desire*, and *joint-intention*.

We distinguish between the individual, group and joint intentions of the agent. The *group-goal* indicates that the agent knows that all team members want to achieve the goal at a time or another. Similarly, *group-desire* and *group-intention* can be defined analogously. For an agent a *group-intention* becomes a *joint-intention* when agents involved in its realisation expressed their mutual belief in this regard, i.e., when the agent knows that this intention is shared by other team members. To form a *joint-intention*, a necessary condition is that the agent must have individual intention to achieve this goal. Similarly, the semantics of joint-desire and joint-goal indicates that all team members have the same *group-desire* and *group-goal* respectively, and all team members know it. Thus, these shared mental attitudes in *task context* of an agent towards the group, specifies that each member holds beliefs about the other team members, and each member mutually believes that every member has the same mental attitude.

To cooperate with other team members, the *joint-intention* is not enough for an agent to engage in the realisation of collective actions. Rather, it only ensures that each member is individually committed to acting. The agent must also ensure the commitment of others to achieve this shared goal. Agents must communicate with other team members to obtain their *joint-commitments*. The agent has a *joint-commitment* towards the group, if and only if, each member of the group has the mutual belief about the same *group-goal*, the agent has the *joint-intention* about to achieve that goal, and each agent of the group is individually committed to achieve this goal. Thus, the shared belief of *task context* also includes the belief about the *joint-commitment* towards group to ensure that every team member has the commitment towards the group to achieve the shared goal. Hence, the IS not only contains information about the current context of the dialogue, but also that of the collaborative task, i.e., beliefs about other team members potentially useful for the agent for its decision-making.

IV. Conversational Behavior

The conversational behavior allows C^2BDI agents to share their knowledge with other team members using natural language communication, and ensures the coordination of the team activity. The agent interprets and generates the dialogues based on the semantics of dialogue acts proposed in [14]. To achieve the coordination among team members, we propose *collaborative conversational protocols* for the agent. These protocols construct the *conversational desires* for the agent which, when activated, result in *conversational intentions*.

A. Collaborative Conversational Protocols

As we want the agent to be proactive and cooperative, we have defined three collaborative conversational protocols (CCPs). These protocols ensure the establishment of the collaboration among team members to achieve the *group-goal*, and its end when the current goal is achieved. Every team member participating in a collaborative activity enters in the collaboration at the same time, and remains committed towards the group until the activity is finished.

a) CCP-1 : When the agent has a new *group-goal* to achieve, it communicates with other team members to establish *joint-commitment*, and to ensure that every team member use the same plan to achieve the *group-goal*.

When the agent has one or more *group-goals* to achieve, and if it has no mutual belief about them, it constructs *Set-Q(what-team-next-goal)* dialogue act and addresses it to the group. By addressing this open question, the agent allows both users and other agents to actively participate in the conversation. If the agent receives the choice of the goal from another team member, it adds a mutual belief about *group-goal* and *group-intention* to its *cognitive context*, and adds the belief about *joint-goal* to the *task context*. It then confirms this choice by sending a positive acknowledgement (by constructing *Auto-feedback(positive-ack)*) to the sender.

When the agent receives *Set-Q(what-team-next-goal)* and has no mutual belief about *group-goal*, i.e., no other team member has already replied to the question, it can decide to reply based on its response time. It chooses one of the available goals based on its own preference rules, and informs sender by constructing *Inform(team-next-goal)* dialogue act. When the agent receives positive acknowledgement from one of the team members, it modifies its IS by adding mutual belief about *group-goal* and *group-intention*, and belief about *joint-goal*.

 ISBN: 978-1-61208-325-4

If the agent has *joint-goal*, but not *joint-intention* to achieve this goal, the agent needs to ensure that every team member will follow the same plan to achieve *group-goal*. If the agent has more than one plan to achieve this goal, it constructs *Choice-Q(which-plan)* act and addresses it to the group, or if the agent has only one plan for the goal, it constructs *Check-Q(action-plan)* act addressing to the group. When the agent receives a choice of the plan, or the confirmation of the choice of a plan, it adds *joint-intention* to its *task context*. It confirms this by sending a positive acknowledgement, and constructs the belief about *joint-commitment*. When the agent receives *Choice-Q(which-plan)* or *Check-Q(action-plan)*, and has no mutual belief about *group-intention*, it constructs *Inform(plan-choice)* or *Confirm* dialogue act respectively to inform about its plan selection. When it receives positive acknowledgement from one of the team members, it adds individual- and joint-commitment to achieve the group-goal.

b) CCP-2 : When the agent has performed all its planned actions of the shared activity, but the activity is not yet finished, the agent requests other team members to inform him when the activity will be finished.

The agent generates *Directive-request(inform-goal-achieved)* to ask other members to inform it when the activity will be finished. When the agent receives this dialogue act, it adds communicative goal *Inform(goal-achieved)* to its agenda.

c) CCP-3 : The agent who finished the last action of the shared activity, informs other team members that the activity is terminated.

The preconditions for CCP-3 are that the agent believes that it has performed the last action of the collaborative activity, and it has the *joint-commitment* to achieve *group-goal*. If these preconditions are satisfied, it constructs *Inform(activity-finished)* dialogue act addressing it to the group. When the agent receives the information that the last action of the activity has been finished, and has the belief about *joint-commitment* in its *task context* and has a communicative goal *Inform(goal-achieved)* to achieve (due to CCP-2), it constructs *Inform(goal-achieved)* dialogue act to inform other team members that the goal has been achieved. It then adds the belief about the achievement of the goal, and removes the corresponding intention from the *task context*. When the agent receives the information about goal achievement, it removes the corresponding intention from the *task context*, and drops the communicative goal *Inform(goal-achieved)* if it has.

The agent waits for certain time (until the threshold of its reaction time is expired) and if no team member has already replied, the agent can create an intention to reply. Otherwise, the agent simply listens to the conversation and updates its beliefs. Thus, in order to establish mutual awareness and to coordinate with other team members, the agent participates in the conversation. Once agents have established the *joint-commitment*, they can coordinate with other team members to achieve the *group-goal*. These protocols are instantiated when the decision-making identifies collaborative situations that satisfy necessary conditions to be fulfilled. These situations add expectations of information from other team members, which need to be satisfied. In a human-agent team, the user's behavior is uncertain, i.e., a user may not necessarily follow these protocols. As the agent updates their beliefs using perception information, which can make expectations to be true from the observation of actions of user perceived by the agent, or from the information provided by other team members. This mechanism makes these protocols robust enough to deal with uncertainty about user's behavior. One of the advantages of these protocols is that the dialogues for the coordination need not to be scripted in the definition of action plans.

B. Decision-Making

In C^2BDI agent, decision-making is governed by the information about current goals, the shared activity plan and knowledge of the agent (IS and semantic knowledge). The decision-making algorithm is shown in Algo. 1. The algorithm

Algorithm 1 DECISION-MAKING ALGORITHM

```
Require: IS
B = IS.SemanticContext.Belief
D = IS.Task-Context.Desire
I = IS.Task-Context.Intention
agenda= IS.Semantic-context.Agenda
while true do
    update-perception(ρ) and Compute B, D, I
    Π ⇐ Plan(P, I )
    while !Π.empty() do
        if agenda is not empty or the agent has received an utterance then
            Process Conversation-Behavior()
            Compute new B, D, I
            Π ⇐ Plan(P, I )
        if the task-focus contains communicative intention then
            Process Conversation-Behavior()
        Identify-Cooperative-Situation in the current shared plan Π
        if Cooperative-Situation is matched then
            Process Conversation-behavior()
        α ⇐ Plan-action(Π), execute(α)
```

verifies whether the agenda in IS is not empty or *task focus* contains communicative intentions. If so, control is passed to the conversational behavior that supports the natural language communication. Otherwise, the agent chooses the plan to be realised. If agent identifies some cooperative situations in the collective activity, where the agent can not progress without assistance, i.e., if the preconditions for one of the CCPs are satisfied, then the control is passed the to the conversational behavior. The cooperative situations generate communicative intentions in the agenda, which causes the agent to interact with other team members to share their knowledge. The agent updates its IS if the control is passed to the conversational behavior, and deliberates the plan to generate the intention. Once the intention is generated, the agent selects actions to be realised and, in turn, updates its *task focus* in IS to maintain the knowledge about the current context of the task.

V. IMPLEMENTATION

This section shows how the C^2BDI architecture has been applied to a collaborative VE for learning of a procedure for industrial maintenance. We illustrate, through a real educational scenario, how decision-making and dialogues allow an agent to coordinate its actions with those of the learner.

A. The Educational Scenario

This scenario describes a maintenance procedure in a plastics manufacturing workshop. The scenario consists in the replacement of a mould in a plastic injection moulding machine (see Fig. 3). This specific intervention requires a precise coordination of tasks between two workers: the setter and the machine operator. The use of autonomous agents allows the learner to execute the learning procedure.

Figure 3: Collaborative realisation of the maintenance procedure in the virtual environment.

Let's consider a situation in which both the user (playing the role of an operator) and the virtual agent (playing the role of a setter) want to replace the mould (see Fig. 4). Following sequence of dialogues describe a typical interaction between them.

A1 : Agent : *What should we do now?* [Set-Q(team-next-goal)]
U1 : User : *We should replace the mould.* [Inform(team-next-goal)]
A2 : Agent : *Ok.* [Auto-feedback(positive-ack)]
A3 : Agent : *Should we use the mould replacement plan?* [Check-Q(action-plan)]
U2 : User : *Yes.* [Auto-feedback(positive-ack)]
.
.
(Agent executes "verify-circuit" action.)
A10 : Agent : *Inform me when you will finish the activity.* [Directive-request(inform-goal-achieved)]
(User executes "lock-the-door" action.)
U8 : User : *What should I do now?* [Set-Q(next-action)]
A11 : Agent : *You have to lock the door.* [Answer(next-action)]
U9 : User : *I have locked the door.* [Inform(action-done)]
A12 : Agent : *We have succeeded to replace the mould.* [Inform(goal-achieved)]

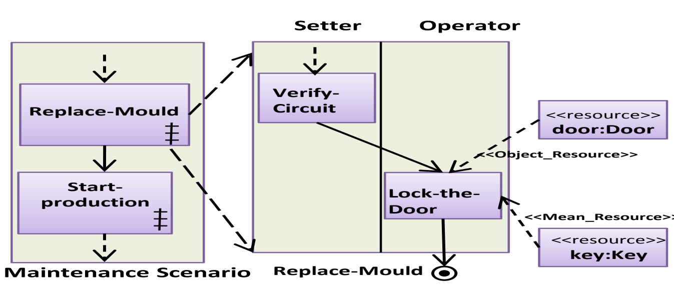


Figure 4: Partial view of the action plan shared between Setter and Operator.

At the beginning, both, the user and the virtual agent have a goal *Replace-Mould*. From the semantic knowledge about the activity, the agent identifies that this goal is shared between team members (in this case, with the user), the goal becomes the group-goal. Table I shows a subset of the agent's knowledge. The agent has a group-goal as Replace-Mould in the IS, but does not have the mutual belief about it. The decision making process identifies this collaborative situation that fulfils conditions of CCP-1 (see Algo. 1, line 15). The CCP-1 generates *Set-Q(team-next-goal)* dialogue act, and adds the communicative intention to the agenda in IS and thus, generates natural language utterance *A1*. When the agent receives utterance *U1*, it interprets *U1* as *Inform(team-next-goal)* dialogue act. As the agent has the same group-goal, it generates positive acknowledgement *A2* for the user and creates mutual-belief about the Replace-Mould (Table II). Now, to ensure that the user will follow the same action plan, the agent constructs *Check-Q(action-plan)* dialogue act considering that the agent has only one plan to achieve group-goal Replace-Mould, and generates *A3*. When the agent receives positive response *U2* from the user, it constructs the joint-intention as well as a joint-commitment to achieve the goal and updates the IS. Now, the decision making process deliberates the plan and computes the new intention (Algo. 1, line 18). Let the current intention of the agent be to *Verify-Circuit*. The subset of agent's knowledge is shown in Table III.

After executing the last action "Verify-Circuit" by the agent from its plan, and as the shared activity is not yet finished, it utters *A10* following CCP-2. The agent interprets the utterance *U8* as an information seeking *Set-Q(next-action)* act, which adds an intention *Answer(next-action)* in its agenda in IS. The decision making process transfers the control to the conversational behavior as the agenda is not empty (see Algo. 1, line 9). By performing the introspection in its shared plan, the agent finds the next action of the user and utters *A11*. Once, user informs the agent that he has finished the last action "lock-the-door" of the shared plan (*U9*), the agent informs that the goal is achieved (*A12*) following CCP-3.

TABLE I: SNAPSHOT OF IS FOR AGENT BEFORE APPLICATION OF CCP-1

	Role R_1 (agent)
Information State	*Task-Context*(group-goal("Replace-Mould"))

TABLE II: SNAPSHOT OF IS FOR AGENT AFTER ESTABLISHING JOINT-GOAL

	Role R_1 (agent)
Information State	*Cognitive-Context*(mutual-belief (group-intention("Replace-Mould") group-goal("Replace-Mould")); *Task-Context*(group-goal("Replace-Mould") joint-goal("Replace-Mould"))

TABLE III: SNAPSHOT OF IS FOR THE AGENT AFTER ESTABLISHING JOINT-COMMITMENT

	Role R_1 (agent)
Information State	*Cognitive-context*(mutual-belief(group-intention("Replace-Mould"); group-goal("Replace-Mould")); *Task-Context*(group-goal("Replace-Mould") joint-goal("Replace-Mould") joint-intention("Replace-Mould") joint-commitment("Replace-Mould") taskFocus(Intention("Verify-Circuit") Intention("Replace-Mould")))

B. Integration with Virtual Agent

The C^2BDI architecture has been integrated with the interaction model for virtual and real human [16] on the GVT platform [17]. The behavior realiser module interacts with the associated virtual agent, and sends requests to it, to perform actions chosen by the decision-making module or by the dialogue manager (turn taking behavior). The user interacts with VE by controlling his avatar thanks to a tracking system of the body and hands. Furthermore, the platform has also been enriched by a voice interface system that uses voice recognition and synthesis of Microsoft (see Fig. 5)].

Figure 5: View of the collaborative scenario with one user.

In C^2BDI architecture, the natural language understanding (NLU) and generation (NLG) is based on the rule based approach [18]. When the agent receives an utterance, it uses NLU rules to determine the corresponding dialogue act type, and the dialogue contents are identified using the semantic knowledge and the contextual information from the IS. The dialogue manager processes these dialogue acts. When the agent has the communicative intention, it constructs dialogue act moves, and NLG rules are used to generate natural language utterances corresponding to the dialogue act.

VI. Conclusion

The proposed behavioral architecture C^2BDI endows the agents in the collaborative VE with the ability to coordinate their activities using natural language communication. This capability allows users and agents to share their knowledge with their team members. The architecture ensures the knowledge sharing between team members by considering the deliberative and the conversation behaviors, not in isolation, but as tightly coupled components, which is a necessary condition for common grounding and mutual awareness to occur. The collaborative conversational protocols we proposed enable agents to exhibit human-like proactive conversational behavior that help users to participate in the collaborative activity. While the implemented scenario already shows the benefits of the solution, the behavior of the agents could be enriched both in terms of collaborative team management and in terms of natural language dialogue modelling. Particularly, it would be interesting to endow agents with problem solving capabilities to select their communicative intentions, or to engage themselves into information seeking behaviors and negotiation rounds, as observed in human teamwork [19].

Acknowledgment

This work was partly supported by the ANR (Corvette project ANR-10-CORD-012).

References

[1] C. Barot, D. Lourdeaux, J.-M. Burkhardt, K. Amokrane, and D. Lenne, "V3S: A virtual environment for risk-management training based on human-activity models," Presence, vol. 22, no. 1, 2013, pp. 1–19.

[2] H. H. Clark and E. F. Schaefer, "Contributing to discourse," Cognitive Science, vol. 13, 1989 , pp. 259–294.

[3] K. Schmidt, "The problem with 'awareness': Introductory remarks on awareness in CSCW," Computer Supported Cooperative Work, vol. 11, no. 3, 2002 pp. 285–298.

[4] X. Fan, J. Yen, and R. A. Volz, "A theoretical framework on proactive information exchange in agent teamwork," Artificial Intelligence, vol. 169, no. 1, Nov. 2005 , pp. 23–97.

[5] P. R. Cohen and H. J. Levesque, "Confirmations and joint action," in Proceedings of IJCAI'91, 1991, pp. 951–957.

[6] B. J. Grosz and S. Kraus, "Collaborative plans for complex group action," Artificial Intelligence, vol. 86, no. 2, 1996, pp. 269 – 357.

[7] C. Rich, C. L. Sidner, and N. Lesh, "Collagen: applying collaborative discourse theory to human-computer interaction," AI Mag., vol. 22, no. 4, Oct. 2001 , pp. 15–25.

[8] J. Bradshaw, P. Feltovich, M. Johnson, L. Bunch, M. Breedy, T. Eskridge, H. Jung, J. Lott, and A. Uszok, "Coordination in human-agent-robot teamwork," in Collaborative Technologies and Systems, 2008. CTS 2008. International Symposium on, 2008, pp. 467–476.

[9] M. Wooldridge and N. R. Jennings, "The cooperative problem-solving process," J. of Logic and Computation, vol. 9, no. 4, 1999, pp. 563–592.

[10] F. Dignum, Dunin-Keplicz, and R. Vebrugge, "Agent theory for team formation by dialogue," in Intelligent Agents VII Agent Theories Architectures and Languages, ser. LNCS. Springer Berlin, 2001, pp. 150–166.

[11] N. Blaylock and J. Allen, "A collaborative problem-solving model of dialogue," in In Proceedings of the SIGdial Workshop on Discourse and Dialog, 2005, pp. 200–211.

[12] K. Kamali, X. Fan, and J. Yen, "Towards a theory for multiparty proactive communication in agent teams." Int. J. Cooperative Inf. Syst., vol. 16, no. 2, 2007, pp. 271–298.

[13] D. Traum and S. Larsson, "The information state approach to dialogue management," in Current and New Directions in Discourse and Dialogue, ser. Text, Speech and Language Technology, J. Kuppevelt and R. Smith, Eds. Springer Netherlands, vol. 22, 2003, pp. 325–353.

[14] H. Bunt, "The semantics of dialogue acts," in Proc. of the 9th Int. Conf. on Computational Semantics, ser. IWCS '11, Stroudsburg, PA, USA, 2011, pp. 1–13.

[15] S. Kopp and N. Pfeiffer-Lessmann, "Functions of speaking and acting: An interaction model for collaborative construction tasks," in D. Heylen, S. Kopp, S. Marsella, C. Pelachaud et H. Vilhjálmsson, editeurs, The First FML workshop, AAMAS, vol. 8, Portugal, 2008.

[16] A. Saraos Luna, V. Gouranton, and B. Arnaldi, "Collaborative Virtual Environments For Training: A Unified Interaction Model For Real Humans And Virtual Humans," in Learning by Playing. Game-based Education System Design and Development, 2012, pp. 1–12.

[17] S. Gerbaud, N. Mollet, F. Ganier, B. Arnaldi, and J. Tisseau, "GVT: a platform to create virtual environments for procedural training," in IEEE Virtual Reality, Reno Etats-Unis, 2008, pp. 225–232.

[18] M. Barange, P. D. Loor, V. Louis, R. Querrec, J. Soler, T.-H. Trinh, E. Maisel, and P. Chevaillier, "Get involved in an interactive virtual tour of brest harbour: Follow the guide and participate," in Proceedings IVA'11, ser. LNCS, vol. 6895. Springer, 2011, pp. 93–99.

[19] W. J. Clancey, "Simulating activities: relating motives, deliberation, and attentive coordination," Cognitive Systems Research, vol. 3, 2002, pp. 471–499.

 ISBN: 978-1-61208-325-4