Subjective Usability of Speech, Touch and Gesture in a Heterogeneous Multi-Display Environment

Arnoud P.J. de Jong Susanne Tak Alexander Toet Sven Schultz Jan-Pieter Wijbenga Jan van Erp
TNO, The Netherlands
{arnoud.dejong, susanne.tak, lex.toet, sven.schultz,jan_pieter.wijbenga, jan.vanerp}@tno.nl

***Abstract*— Several interaction techniques have been proposed to enable transfer of information between different displays in heterogeneous multi-display environments. However, it is not clear whether subjective user preference for these different techniques depends on the nature of the displays between which information is transferred. We explore subjective usability of speech, touch and gesture for moving information between various displays in a heterogeneous multi-display environment, consisting of a multi-touch table, a wall-mounted display and a smartphone. We find that subjective user evaluation of the various interaction techniques depends on the combination of displays being used. This implies that the type of display combination should be taken into consideration when designing interaction techniques for the transfer of items between displays in a heterogeneous multi-display environment. Also, gesture based interactions were judged more acceptable when they involved holding a mobile phone, probably since this provided a cue explaining the action.**

Keywords- large display; multi-display environment; multi-touch table; smartphone; speech; gestures

I. INTRODUCTION

Distributed computing environments (e.g. meeting rooms, collaborative work spaces) are increasingly populated with many heterogeneous display devices like smartphones and tablets (providing small personalized displays), tabletop displays (facilitating collaboration between small groups), and large size displays (for information presentation to larger groups).

A frequent task in these heterogeneous multi-display environments is moving objects between displays [1]. A typical example is the exchange of files (e.g. images or documents), either from a mobile device to a tabletop (for group activities), from a tabletop or mobile device to a large wall display (for public presentation), or from a large screen or tabletop to a mobile device (for personal use).

To minimize user error and workflow interruptions the techniques for cross-display interaction should be simple and intuitive, requiring minimal physical and cognitive effort. Also, they should be "socially acceptable" (i.e. they should not make the user feel uncomfortable).

Although several cross-display interaction techniques have been proposed, it is still unknown if subjective user preference for these different techniques depends on the nature of the selected display pairs. Therefore, the current study assesses subjective user preferences for gesture, touch and speech based interaction techniques in a setting that integrates a multi-touch table with a wall-mounted display and a smartphone. To the best of our knowledge this is the first study to investigate these three interaction modalities in the same heterogeneous multi-device collaborative computing environment. In the rest of this paper, we will first discuss related work on interaction and navigation techniques used in multi-display environments. Then we will describe a user study that we performed to assess the subjective usability of gesture, touch and speech techniques for this purpose, and we will present the results of this study. Finally, we will present our conclusions and we will provide suggestions for future research.

II. RELATED WORK

Multiple or distributed display environments present new challenges for interaction and navigation. Currently the interaction with heterogeneous multi-display environments is still dominated by a single-user single-display paradigm: the user interacts with one display at a time, using an interaction technique that is considered most appropriate for a particular type of display. Available interaction techniques are typically keyboard, touch, gesture, or speech based. Keyboard input has long been the standard but can be too slow and cumbersome in dynamic environments. Touch based interaction has become a popular interaction technique for devices like mobile phones, tablets and interactive tabletops. Although it is generally fast, it is only suitable for direct interaction at close range. Gestural interaction has gained popularity since the application in interactive computer games (Wii, Microsoft's Kinect system). This technique is more appropriate for direct interaction with large displays that can be operated from a distance [2]. Speech based interaction has become a common direct interaction technique for in-car navigation devices and hands-free phone systems, and might gain in popularity with the increasing availability of voice operated smartphone apps (e.g. via Siri on the iPhone). Similar to touch, speech interaction is only suitable for direct interaction at close range.

Recently several new interaction techniques have been proposed to move objects between tabletop displays and mobile phones [3-6], tablets [1, 7] or hand held devices in general [8], between hand-held and large displays [9, 10], or between any of these devices [11]. Many of these techniques are gesture based.

Speech and gestures complement each other and (when used together) can create an interaction technique that is more powerful than either modality alone. Speech interaction is suited for descriptive techniques, while gestural interaction

 ISBN: 978-1-61208-250-9

is ideal for direct manipulation of objects [12]. Speech allows interaction with objects regardless of their degree of visual exposure (occlusion). It appears that users prefer using combined speech and gestural interaction over either modality alone when handling graphics manipulation [13].

III. USER STUDY

To assess user preference for different interaction techniques in a heterogeneous multi-display environment, we performed a study in which users transferred items (photographs) between different types of displays, using gesture, touch and speech techniques. Subjective user experience was quantified through semantic questionnaires.

A. Method

A setup with a multi-touch table, a wall-mounted screen and a smartphone was used (see Figure 1). Participants performed the same task (i.e. transferring a photograph from one display to another) several times for four different display pairs, using various techniques.

B. Task and interaction techniques

Participants were requested to send a photograph from one display to another using various techniques. The four display pairs were (touch)table to screen, table to mobile (phone), screen to mobile, and mobile to screen. The interaction techniques were as follows:

Table to screen
Speech (A1): Select photo and say "*send to screen*".
Touch (A2): Drag photo to a window entitled '*Screen'*.
Gesture (A3): Select photo and point at the screen.

Table to mobile
Speech (B1): Select photo and say "*send to Harry*".
Touch (B2): Drag photo to a window entitled '*Harry'*.
Tangible (B3): Place mobile on table and drag photo to it.

Screen to mobile
Speech (C1): Start voice command by saying "*screen*", then say "*send to Harry*".
Gesture (C2): Hold phone as if taking a photo of the screen.

Mobile to screen
Speech (D1):Start voice command by dragging finger downwards over the screen and say "*send to screen*".
Touch (D2): Press send button below photo and select the '*Screen'* menu item.
Gesture (D3): Point phone at the screen.

C. Design and procedure

Participants completed tasks in all display pair × technique combinations, using a repeated measure within-subject design. The four different display pairs were presented in random order. For each display pair, the experimenter first demonstrated the various interaction techniques. To familiarize the participants with the interaction techniques they were given the opportunity to practice. Next, participants were requested to perform the interaction (transferring the photo from one display to the other) five consecutive times for each interaction technique. After completing the tasks for a particular interaction technique a questionnaire was administered verbally. The questionnaire contained seven statements, each related to a particular aspect of perceived usability:

1. I could execute the task without thinking (*without thinking*);
2. The interaction was intuitive (*intuitive*);
3. The interaction felt unnatural (*unnatural*);
4. The interaction was tiring (*tiring*);
5. The system responded quickly (*responsive*);
6. The interaction is complex (*complex*);
7. The interaction is error-prone (*error-prone*).

Participants rated their agreement on a 5-point Likert scale ranging from "completely disagree" to "complete agree". In addition, for each display pair, participants were asked to rank the three (two for the screen to mobile display pair) techniques from most to least preferred. On average, participants completed the experiment in 60 minutes.

D. Participants

Twenty-one people participated in the user study (12 male, 9 female, 20-57 years old, average age 27). Fourteen participants owned or regularly used a device that uses touch input (e.g., smartphone or tablet). The experiment was undertaken with the consent of each participant. Participants were paid 30 euro for their participation.

IV. RESULTS

The questionnaire results show significantly different ratings for all questions (Friedman test, $p < .001$ for *without thinking*, *intuitive*, *unnatural*, *complex* and *error-prone*, $p < .01$ for *tiring* and *responsive*). Post hoc pairwise comparisons (Wilcoxon with Bonferroni correction, $\alpha = .05$) were used to examine the questionnaire results in more detail. Where relevant, these results are discussed in the sections on each of the display pairs below.

The seven questions were converted to one overall subjective usability score by converting all answers to scores ranging from -2 to 2: completely agreeing on a positively framed statement such as without thinking was scored 2, while the reverse coding scheme was applied to negatively framed statement such as *unnatural*. Next, these scores were summed and averaged for all conditions (mean Cronbach's alpha .76), leading to overall scores in the range [-2,2]. The results are shown in Figure 2.

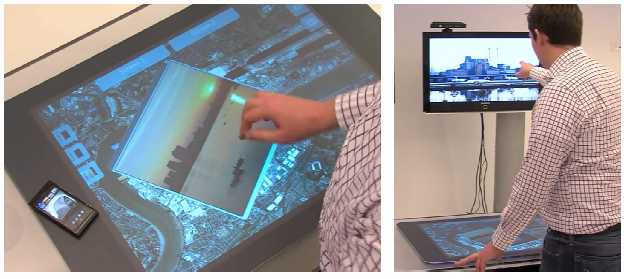

Figure 1. Experimental setup; interacting with the multi-touch table (left) and the screen (right)

 ISBN: 978-1-61208-250-9

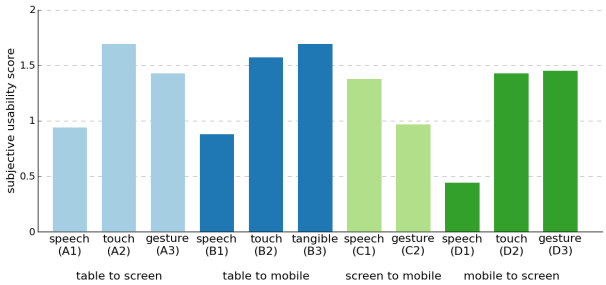

Figure 2. Subjective usability scores for the 11 experimental conditions.

The subjective usability scores (measured across the 11 different conditions) were analyzed using a one-way repeated-measures ANOVA [14]. The results show a main effect for condition ($F_{10,200}=11.1$, $p<.001$), which was further analyzed using post-hoc pair-wise comparisons with Bonferroni correction ($\alpha=.05$). These results are discussed in the sections on each of the display pairs below.

A. *Table to Screen*

Analysis of the subjective usability scores revealed significant differences between the speech technique (A1) and the touch technique (A2) ($p<.05$). Participants had a clear preference for the touch technique; 17 participants preferred touch the most, 4 participants preferred gesture the most, and no participants preferred speech the most (Figure 3, $X2(2)=22.57$, $p<.001$). Further analysis of the qualitative remarks by the participants revealed that 5 participants explicitly reported that they thought speaking the commands out loud was awkward. In general, people were very positive about the touch technique. Finally, the gesture technique was also well received by the participants, with 4 participants reporting that the method was fun to use. However, 4 other participants remarked that they considered this particular gesture (stretching their arm and pointing) to be embarrassing or too "commandeering".

B. *Table to mobile*

Analysis of the subjective usability scores revealed significant differences between the speech technique (B1) and the touch technique (B2) ($p<.05$), and the speech technique (B1) and the tangible technique (B3) ($p<.01$). This is also reflected in the preferences for the various techniques; 15 participants preferred the tangible technique the most, 5 participants preferred gesture the most and 1 participant preferred speech the most ($X2(2)=14.86$, $p<.001$). Furthermore, 18 participants preferred speech the least (Figure 3). In particular, the results showed that people considered the speech technique for this task type significantly more unnatural than both the touch and the tangible techniques (both $p<.05$), which was also evident from the qualitative remarks made by the participants (similar to the table to screen task, 4 participants reported feeling awkward when speaking the commands out loud). The tangible technique was very well received, with 10 participants calling it "fun" or "cool".

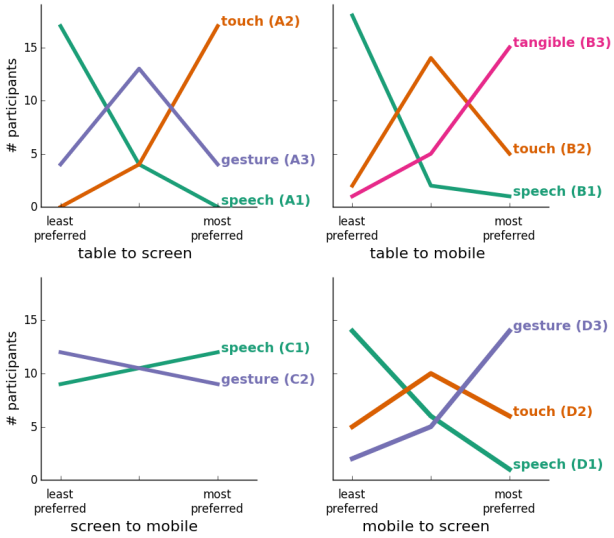

Figure 3. Number of participants preferring a certain interaction mode (touch, gesture, tangible, speech) for each of the 4 different device pairs.

C. *Screen to mobile*

Analysis of the subjective usability scores revealed no significant difference between the speech technique (C1) and the gesture technique (C2). There was also no significant preference for either of the techniques; 12 participants preferred speech the most, and 9 participants preferred gesture the most ($X2(1)=0.2$, $p>.6$). This is in contrast to the other display pairs, where speech was generally the least preferred technique (Figure 3). Analysis of the qualitative remarks made by the participants reveals that participants had no reservations about "talking to the screen". This is in contrast to the table to screen/mobile tasks, where people felt uncomfortable using speech.

D. *Mobile to screen*

Analysis of the subjective usability scores revealed significant differences between the speech (D1) and touch technique (D2) ($p<.05$), and between the speech (D1) and gesture technique (D3) ($p<.01$). This is reflected in the user preferences: 14 participants preferred the gesture technique the most, 6 participants preferred touch the most, and 1 participant preferred speech the most ($X2(2)=12.29$, $p<.01$).

V. CONCLUSIONS AND FUTURE WORK

Overall, the results show that subjective user preference for the interaction techniques depends on the type of task.

In general, the speech technique was not very well liked: people often reported feeling embarrassed when speaking commands out loud. Note that the voice commands were in English, which probably introduced an extra degree of difficulty for the participants, who were not native English speakers. In addition, some participants had strong preconceptions about speech being an inappropriate

 ISBN: 978-1-61208-250-9

interaction mode. Speech was only the preferred interaction mode when the large screen was the target of the speech command (C1; see Figure 2). Possibly, speech was deemed acceptable in this case because there are fewer viable interaction alternatives (the screen is too distant to touch).

Though gesture-based interaction techniques were generally well received (except for some participants having issues with the required physical effort), there are some interesting differences between the different tasks. In particular, the use of gestures was not preferred in the table to screen task (A3), while it was positively received on the mobile to screen task (D3, see Figure 3). This is interesting because the gestures are similar for both task types, with the key difference being whether the participant is holding an object (smartphone) or not. Participants regularly felt uncomfortable or embarrassed using a gesture for the table to screen task, but not for the mobile to screen task. Possibly, holding an object that provides a clear visual cue explaining the user's actions makes gesture-based interaction more acceptable [15, 16]. Gesture-based interaction techniques may become more acceptable when people are allowed to use tangible objects, perhaps even if these objects serve no technical purpose (i.e. a dummy object). We note that the subjective usability scores showed no difference between the various gesture conditions, but the user preference rankings did (see Figures 2 and 3). Also, qualitative remarks made by the participants suggest a difference between the various gesture conditions. This suggests that our questionnaire was incomplete in that sense, and that future user evaluations should explicitly address embarrassment. More specifically, in future studies we intend to test whether participants are primarily spatial, verbal or object oriented.

This study focused on the subjective evaluation of different interaction techniques. In future studies we also intend to register objective performance measures (e.g. the time it takes to perform the different actions).

Finally, we note that we only investigated single users interacting with single display pair, with only the experimenter present. Future research could investigate settings with multiple users and/or more complex display combinations.

REFERENCES

[1] Nacenta, M.A., Aliakseyeu, D., Subramanian, S. & Gutwin, C. (2005). A comparison of techniques for multi-display reaching.Proc. CHI '05 (pp. 371-380). New York, USA: ACM.

[2] Bragdon, A. & Ko, H.-S. (2011). Gesture Select: acquiring remote targets on large displays without pointing.Proc. CHI '11 (pp. 187-196). New York, USA: ACM Press.

[3] Chehimi, F. & Rukzio, E. (2010). Throw your photos: an intuitive approach for sharing between mobile phones and interactive tables.Proc. Ubicomp 2010 New York, USA: ACM Press.

[4] Shirazi, A.S., Döring, T., Parvahan, P., Ahrens, B. & Schmidt, A. (2009). Poker Surface: combining a multi-touch table and mobile phones in interactive card games.Proc. MobileHCI '09 New York, USA: ACM P.

[5] Schmidt, D., Chemini, F., Rukzio, E. & Gellersen, H. (2010). PhoneTouch: a technique for direct phone interaction on surfaces.Proc. ITS '10 New York, NY, USA: ACM.

[6] Yoo, J.-W., Hwang, W., Seok, H., Park, S.K., Kim, C., Choi, W.H. & Park, K. (2010). Cocktail: exploiting bartenders' gestures for mobile interaction. Int.J.Mob.Hum.Comp.Interaction, 2(3), 44-57.

[7] Bader, T., Heck, A. & Beyerer, J. (2010). Lift-and-drop: Crossing boundaries in a multi-display environment by Airlift.Proc. AVI'10 (pp. 139-146). New York, USA: ACM Press.

[8] Olsen, D.R., Clement, J. & Pace, A. (2007). Spilling: expanding hand held interaction to touch table displays.Proc. TABLETOP '07 (pp. 163-170). Los Alamitos, CA, USA: IEEE Press.

[9] Hutama, W., Song, P., Fu, C.-W. & Goh, W.B. (2011). Distinguishing multiple smart-phone interactions on a multi-touch wall display using tilt correlation.Proc.CHI '11 (pp. 3315-3318). New York, USA: ACM Press.

[10] Jeon, S., Hwang, J., Kim, G.J. & Billinghurst, M. (2006). Interaction techniques in large display environments using hand-held devices.Proc. VRST '06 New York, NY, USA: ACM Press.

[11] Kray, C., Nesbitt, D., Dawson, J. & Rohs, M. (2010). User-defined gestures for connecting mobile phones, public displays, and tabletops.Proc. MobileHCI '10 (pp. 239-248). New York, NY, USA: ACM Press.

[12] Oviatt, S. (1999). Ten myths of multimodal interaction. Comm.ACM, 42(9), 74-81.

[13] Hauptmann, A.G. & McAvinney, P. (1993). Gestures with speech for graphic manipulation. Int.J.Man-Machine Studies, 38(2), 231-249.

[14] [14] Carifio, J. & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. Medical Education, 42(12), 1150-1152.

[15] Rico, J. & Brewster, S. (2010). Usable gestures for mobile interfaces: Evaluating social acceptability.Proc. CHI '10 New York, NY: ACM Press.

[16] Montero, C.S., Alexander, J., Marshall, M.T. & Subramanian, S. (2010). Would you do that?: Understanding social acceptance of gestural interfaces.Proc. MobileHCI '10 New York, USA: ACM Press.

 ISBN: 978-1-61208-250-9