

Elckerlyc Goes Mobile

Enabling Natural Interaction in Mobile User Interfaces

Randy Klaassen, Jordi Hendrix, Dennis Reidsma,
Rieks op den Akker, Betsy van Dijk
Human Media Interaction
University of Twente
Enschede, the Netherlands
{*r.klaassen, j.hendrix, d.reidsma,*
h.j.a.opdenakker, e.m.a.g.vandijk}@utwente.nl

Harm op den Akker
Roessingh Research and Development, Telemedicine group
University of Twente, Telemedicine group
Enschede, the Netherlands
h.opdenakker@rrd.nl

Abstract—The fast growth of computational resources and speech technology available on mobile devices makes it possible to entertain users of these devices in having a natural dialogue with service systems. These systems are sometimes perceived as social agents and this can be supported by presenting them on the interface by means of an animated embodied conversational agent. To take the full advantage of the power of embodied conversational agents in service systems it is important to support real-time, online and responsive interaction with the system through the embodied conversational agent. The design of responsive animated conversational agents is a daunting task. Elckerlyc is a model-based platform for the specification and animation of synchronised multi-modal responsive animated agents. This paper presents a new light-weight PictureEngine that allows to run this platform in mobile applications. We describe the integration of the PictureEngine in the user interface of two different coaching applications and discuss the findings from user evaluations. We also conducted a study to evaluate an editing tool for the specification of the agent's communicative behaviour. Twenty one participants had to specify the behaviour of an embodied conversational agent using the PictureEngine. We may conclude that this new light-weight back-end engine for the Elckerlyc platform makes it easier to build embodied conversational interfaces for mobile devices.

Keywords- *Mobile ECA; User Interfaces; user evaluations; (mobile) coaching applications.*

I. INTRODUCTION

Advances in user interface technology — speech recognition, speech synthesis and screen capacities — more and more allow people to engage in spoken interaction with services on their mobile phones. The use of a talking head or an embodied conversational agent (ECA) can support spoken interaction in different kind of user interfaces. In [1], a new light-weight PictureEngine was presented that allows to use ECAs in the user interfaces of mobile applications. The presentation of a service agent by means of a persona supports the idea of the computer as a social actor. Research has shown that animation of human-like social behaviours and expressions by means of a virtual human or embodied conversational agent strengthens the impression that the

agent is present and engaged in the interaction [2]. They have a positive effect on user experience [3].

In human-human conversations, the one who has the speaker role is monitoring his addressees while speaking. Listeners give back-channels, short comments, and may also interrupt the speaker. By his gaze behaviour the speaker shows his interest with the addressee. By adjusting or stopping his speech, he shows being responsive to the listeners comments and that he is really engaged in the conversation. Gaze behaviour in conversations is important for interaction management, in particular for signalling that one wants to have the floor, that the speaker wants to keep the floor or is willing to yield the floor. Emotion expressions are prime indicators of engagement in what is going on in the conversation [4]. In designing virtual humans that are able to show these social signals and responsiveness one needs well designed model-based specification languages and tools. The SAIBA framework [5] provides a good starting point for designing the behaviours of interactive virtual humans. Its Behaviour Markup Language (BML) defines a specification of the form and relative timing of the behaviour (such as speech, facial expression, gesture) that a BML realizer should display on the embodiment of a virtual human. Elckerlyc is a state-of-the-art BML Realizer. In [6], its mixed dynamics capabilities are described and its focus on continuous interaction, which makes it very suitable for virtual human applications requiring high responsiveness to the behaviour of the user.

The Elckerlyc platform can act as a back-end realizer for different embodiments, like physical robots or realistic 3D full kinematic virtual humans. Using a full 3D virtual human on a mobile phone is however too heavy in terms of processing power and battery usage. To be able to use the Elckerlyc platform on a mobile phone, a light-weight animation embodiment is needed. The work presented here contributes to satisfy this need.

One of the many application areas for natural interaction with embodied agents is in healthcare services and coaching

systems that users interact with through mobile devices. The field of Telemedicine — healthcare delivered remotely to the patient or user — is receiving a large amount of attention as a promising paradigm to reduce the burden on traditional healthcare services. As the population in the Western world is ageing, the prevalence of chronic disease is rising and the cost of healthcare increasing. Technology aided coaching on healthy behaviour (such as daily physical activity) can help prevent chronic diseases and influence the process of healthy ageing in general. Activity coaching is also potentially useful for everyone. The American College of Sports Medicine recommends in its 2011 position paper that every healthy adult engages in at least 30 minutes of moderate physical activity for five days per week [7]. Smart phones offer a unique opportunity to deliver coaching on physical activity to the user, as they become increasingly ubiquitous and are capable of running increasingly complex applications. They also contain built-in sensors, enabling context aware intelligent coaching through the use of location- and web based services. Research by, e.g., Bickmore [3] showed that personification of the user interface of coaching systems can have positive effects on the effectiveness of the coaching. Some examples of personification of the user interface of mobile coaching applications can be found in [8][9][10][11]. Some of these systems are distributed systems, while other system do not display (real-time) animations.

This paper presents the PictureEngine, a light-weight animation embodiment that enables our SAIBA-based BML realizer to be implemented and run stand alone on mobile Android devices. Compared to static pictures or pre-recorded movies, real-time animations are able to react immediately to the user, and this responsiveness increases the experience of engagement of the agent. Section II describes the Elckerlyc platform in more detail. The PictureEngine will be discussed in Section III, and the implementation of the platform on Android in Section IV. Section V presents the results of an evaluation of a design tool that helps designers in specifying their own ECA behaviour using the Behaviour Markup Language. This tool uses the PictureEngine to implement the multi-modal interactive behaviour specification. In Section VI two applications are discussed in which the Picture-Engine was integrated as part of the user interface. These applications are context-aware physical activity coaching applications. The ECA developed for these applications presents feedback of the digital coach by animated spoken interaction. We present some small user evaluation studies in which these coaching systems are evaluated. We conclude with describing future work on the development of the mobile embodied coach.

II. THE ELCKERLYC PLATFORM

In behaviour generation, at least two main aspects can be distinguished. The first aspect is the planning of the actions and movements as means to a certain goal that

the agent intends to achieve. The second one is the actual detailed realization of the verbal and non-verbal behaviours in terms of “embodiments” of the (graphical) virtual human — including the generation of the speech by a text-to-speech synthesizer. This distinction between intent planning, behaviour planning and behaviour realization is the basis of the SAIBA framework (see [12] for more information about the SAIBA framework) [13]. According to this framework the detailed behaviours are specified in the Behaviour Markup Language [14].

The Elckerlyc platform is a BML realizer for real-time generation of behaviours of virtual humans (VHs). The Elckerlyc platform has been described and compared with other BML behaviour realizers (for example EMBR [15] and Greta [16]) in various papers [6][17][18].

Depending on the application and task that the intelligent system has, the virtual human presents for example the character of a tutor, an information assistant, or a conductor [19]. The goal is to make these embodied conversational agents look like believable and convincing communicative partners while interacting with humans. This requires the generation and coordination of “natural” behaviours and expressions.

Reidsma and Welbergen [18] discuss several features of the modular architecture of Elckerlyc and relates each of them to a number of user requirements. A general overview of the Elckerlyc system can be found in Figure 1. The input of the Elckerlyc platform is a BML specification. BML provides abstract behaviour elements to steer the behaviour of a virtual human. A BML realizer is free to make its own choices concerning how these abstract behaviours will be displayed on the embodiment. For example, in Elckerlyc, an abstract ‘beat gesture’ is by default mapped to a procedural animation from the Greta repertoire [20]. Greta is an expressive ECA that is able to show complex emotions. The developer may want to map the same abstract behaviour to a different form, i.e., to a high quality motion captured gesture [18]. The Elckerlyc system is easy to extend with other engines. Different Engines will handle their own parts of the behaviour specification and generate synchronized instructions for realizing, i.e., speech output, body gestures, postures and facial expressions. The output of all the engines are displayed on one embodiment, like a realistic 3D full kinematic virtual human, the Nabaztag or a graphical 2D cartoon like picture animation. Figure 2 shows three examples of embodiments supported by the Elckerlyc platform.

Not every embodiment is able to render all the behaviours that can be specified in BML. This depends on what the embodiment offers, i.e., a robot that is not able to smile or a picture animation that lacks a picture showing the smiling face cannot render the requested smiling behaviour. The interface between the output of Elckerlyc and the embodiment occurs in a Binding. A Binding is an XML

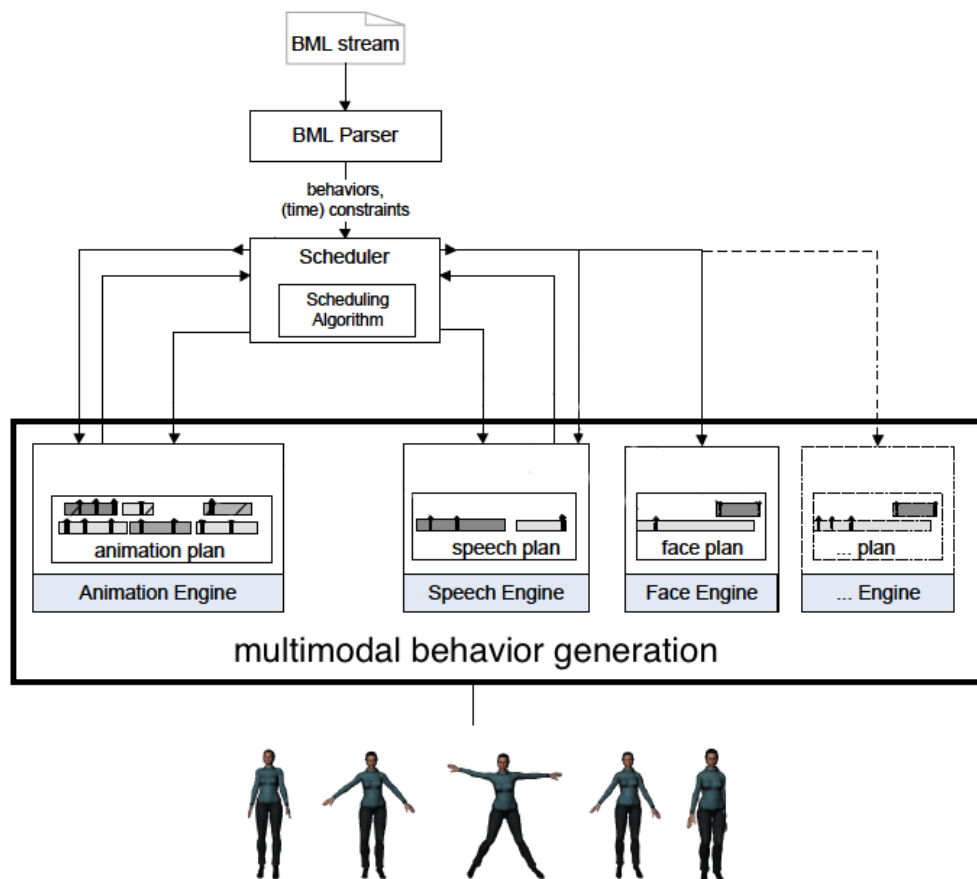


Figure 1. Overview of the Elckerlyc architecture. BML input is processed by the Elckerlyc system by different engines. The result is combined into one embodiment. New Engines, or Engines developed by others are indicated by dash lines.

description to achieve a mapping from abstract BML behaviours to PlanUnits that determine how the behaviour will be displayed in the embodiment. Bindings can be customized by the application developer. Other Engines provide similar bindings.

This paper discusses how this feature was exploited. A light-weight PictureEngine was developed that makes it possible to run Elckerlyc on mobile Android platforms. Elckerlyc allows for a transparent and adjustable mapping from BML to output behaviours (rather than the mostly hard coded mappings in other realizers), and allows for easy integration of new modalities and embodiments, for example to control robotic embodiments, or full 3D embodiments. The PictureEngine that was developed allows rendering of behaviours and expressions using layers of pictures. The next section will discuss the PictureEngine in more detail.

III. THE PICTUREENGINE

Using a realistic 3D full kinematic virtual human embodiment is not suitable for use on mobile devices for multiple

reasons. Not only do such devices lack the processing power to render this kind of environment, but displaying a full scene including a full body ECA on the relatively small screen of a mobile device is quite impractical. The displayed size of the ECA would make it so small that its expressions would hardly be visible. The high processing demands would also drain the devices' battery quickly. In order to avoid all these problems, Elckerlyc uses a different graphical embodiment on the Android platform, the PictureEngine.

The PictureEngine is a lightweight graphical embodiment that uses a collection of 2D images in order to display the ECA. While having a 2D image embodiment offers some limitations, it also has its advantages. First of all, it has low demands in terms of processing and memory. It also allows for great variation in the design of ECA's. One could for example design a cartoon figure ECA, an ECA based on more lifelike illustrations, or even an ECA based on photographic images of a real person, or pre-rendered images of a 3D lifelike ECA. Creating your own 2D ECA starts with designing a set of images for the appearance of

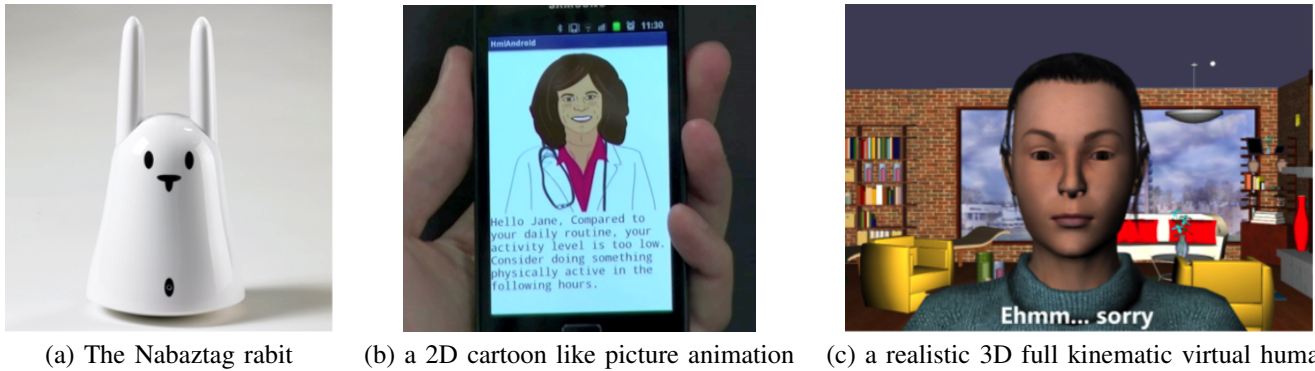


Figure 2. Three types of embodiment used as back-end for the Elckerlyc platform

the ECA in different layers. More detailed information about layers can be found in Section III-A. Blinking, lip-sync and other nonverbal behaviour can be designed as a set of animations. Animations are designed by a subset of pictures that together can execute the animation and are defined in a small XML file. Section III-B will explain the animations in more detail. When these different images and animations are designed they have to be specified in a PictureBinding which will be explained in Section III-C.

A. Layers

In order to generate a dynamic ECA from a collection of images, the PictureEngine uses a layer-based approach. Different parts of the ECA are displayed on different layers of the final image, and can thus be in different states. For example, one layer may contain the eyes, while another contains the mouth. Figure 3 shows an example of how a couple of layers will result in a face of the ECA. The base layer normally contains the ECA in a base state, meaning that when the ECA is in a neutral or passive state, the user sees only this base layer. While each (facial) feature of the ECA does have its own layer, they are also present in the base layer. The base layer contains for example a full face with a neutral expression, even though the eyes and mouth may have their own layers. There can also be layers containing features that are not visible in the base state, such as hands that only move into the frame when executing a gesture. By using this layered approach, different parts of the ECA can be manipulated independently and combined in order to generate different expressions. This also allows the ECA to do several (connected or unconnected) things at once, such as blink while also speaking and pointing at something.

The layered approach does present some limitations. Since the features of the ECA are in separate layers, the base onto which these features are displayed (usually a face, and possibly part of the body) is generally static, so any movement of the entire ECA poses a problem. When an ECA contains facial features on different layers, the layered structure prevents it from moving around. This also applies

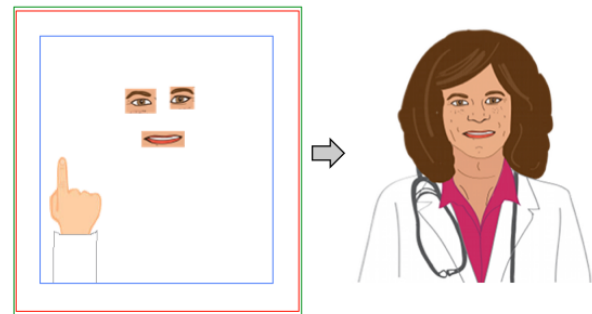


Figure 3. An example of different layers that finally results in the face of a ECA.

to smaller movements such as nodding, shaking and tilting of the head. However, because the PictureEngine is designed to be used on smaller screens, the ECA will generally be displayed as a talking head, a close-up of a face covering most of the available screen space. Subsequently, having the ECA perform locomotion is already impractical and, since there is hardly any room for the ECA's environment to contain anything but itself, arguably unnecessary.

B. Animations

While single images may suffice for portraying expressions in many cases, there are other cases where an ECA simply has to display some motion in order to come across as believable. To make this possible, the PictureEngine also allows the use of animations instead of single images. These animations are defined by using a simple XML format that allows a number of images to be listed, together with the duration for which they are to be displayed. While these durations are specified in seconds, the nature of the BML scheduler allows the duration of animations to be adjusted according to the BML code that is being realized, causing the animation to play faster or slower depending on the timespan determined by the scheduler.

An additional feature of these animation XML files that provides an advantage over using an already established

```

<PictureUnitSpec type="face">
  <constraints>
    <constraint name="type" value="LEXICALIZED"/>
    <constraint name="lexeme" value="smile"/>
  </constraints>
  <parametermap>
  </parametermap>
  <parameterdefaults>
    <parameterdefault name="filePath" value="animations/" />
    <parameterdefault name="fileName" value="smile.xml" />
    <parameterdefault name="layer" value="8" />
  </parameterdefaults>
  <PictureUnit type="AddAnimationXMLPU"/>
</PictureUnitSpec>

```

Figure 4. PictureBinding entry for a smile.

format for image animations is the possibility to include synchronization information in the animation specification. Between any two frames of an animation, a synchronization point can be included in the specification. These synchronization points are available for use in the main BML code. In this way, it is possible to, e.g., synchronize the stroke of a beat gesture animation with a certain word within a speech element.

C. PictureBinding

In the same way that the other engines uses bindings, the PictureEngine uses a PictureBinding. This PictureBinding allows a combination of a BML behaviour class and (optionally) several constraints to be mapped to a certain image or animation. It is possible to include anywhere from zero constraints to all the constraints defined by the corresponding BML behaviour type. This allows the designer of a PictureEngine ECA to refine those behaviours that are most relevant to the ECA, and implement any others in a more general fashion.

The actual PictureBinding itself is defined in an XML file containing the behaviour classes and constraints and the PictureUnits and parameters they are to be mapped onto (see Figure 4 for an example). The accessibility of this format allows an ECA to be designed or modified by someone who does not have knowledge of the inner workings of Elckerlyc. Only knowledge of BML and the available PictureUnits and their parameters is required to be able to build a complete PictureBinding.

D. Lip-sync

In order to visually display the fact that the ECA is speaking, the PictureEngine provides a rudimentary lip-sync facility. This lip-sync feature is implemented in the same way as the lip-sync provided by the AnimationEngine. However, where the AnimationEngine provides a full mapping from visemes to animation units, the PictureEngine lip-sync currently does not make use of such a mapping (although it could be added in the future). In its current state the lip-sync allows a single animation to be specified which is played whenever the ECA is speaking. This animation is repeated

for the number of times it fits into the duration of the speech unit (and slightly adjusted so that the amount of repetitions becomes a round number).

IV. ANDROID IMPLEMENTATION

Since the Elckerlyc platform is implemented almost entirely in Java, all of its core elements run on Android without any modification. However, since Android has its own environment for visual and audio output, some additions are required. This does not mean that the Android application uses a modified version of the core Elckerlyc platform. The fact that Elckerlyc uses an XML format to define the loading requirements for a specific ECA allows the Android application to simply load its own versions of a few key components. This allows the core Elckerlyc system to be used in the Android application as-is, so any changes to the Elckerlyc core can be directly used in the Android application without having to modify or port it first. The Android application requires Android Gingerbread (2.3) or higher. The subsystems for which the Android application contains its own versions are discussed in this section.

A. Graphical Output

The Android platform has its own graphical environment. Therefore the engines that provide graphical output use a modified component for printing this output in the Android application. This goes for both the PictureEngine, which handles the graphical display of the ECA, and the TextSpeechEngine, which outputs speech elements to a text area. Since PNG images can be handled without problems by the Android graphical environment, the additional code needed to replace the PictureEngines' default output subsystem with a version that works on Android is minimal. Displaying plain text is a basic function in Android.

B. SpeechEngine

In the case of the SpeechEngine (for the rendering of spoken text using text-to-speech(TTS)) the differences with Android are unfortunately more severe. The TTS engines used in the PC version of the SpeechEngine contain several dependencies on native PC systems and cannot be used on Android without significant changes. However Android does offer an internal TTS system. Using this internal system avoids the costly process of porting a TTS engine and any possible efficiency issues this may bring. In order to make use of the internal Android TTS system, an Android adaptation of the Elckerlyc SpeechEngine is needed. This includes the module that loads and initializes the engine, as well as the parts of the system dealing with the actual TTS operations.

The main problem with the Android TTS system is that it is not possible to obtain timing information for utterances, meaning there is no way to find out exactly at what time a word is spoken. This causes the BML scheduler to be

unable to use synchronization points within utterances. This makes it hard to precisely synchronize other behaviours with specific words being spoken. A partial solution is that utterances are pre synthesized to a file in order to find the total duration of the utterance. This provides the crucial information for the Elckerlyc scheduler. This “preloading” of utterances causes a delay at start up before the ECA starts playback of the requested BML code. Furthermore, the TTS also does not offer any viseme information, making it impossible to use real lip-sync on Android. This is the main reason the PictureEngine on Android does not currently support true lip-sync.

C. Subtitles

Because the PictureEngine can run on a mobile device, the chances of the user having trouble hearing the text spoken by the TTS on the Android system are quite high. This could be caused by factors such as environment noise, low volume or bad speakers. In order for the user to still be able to interact with the ECA in these situations, the Android application also offers an on-screen representation of any spoken text, comparable to subtitling. The TextSpeechEngine (on-screen text display) receive the text handled by the SpeechEngine and displays spoken text to in a text area, synchronized (per utterance) with the TTS.

V. USER EVALUATION WITH THE PICTUREENGINE

This section presents a user evaluation of a BML editor that uses the PictureEngine. The goal of this user evaluation was to investigate if it was possible for non expert users to specify the verbal and non-verbal behaviour of an ECA in BML. Users had to create their own BML script following a step-by-step task description. This results in verbal and non-verbal behaviour of an ECA which is displayed by means of the PictureEngine. With this evaluation we tried to find out what problems users meet when creating a BML script. We also wanted to know how users assessed the usability of the BML editing tool.

A. Procedure

Participants were students and employees of the University of Twente as well as partners from the Smarcos project. Participants had to simply download and install the BML editor including the PictureEngine. An editor screen is used to edit the BML script and a feedback screen shows feedback and error messages (see Figure 5). On a separate screen to the ECA is displayed (see Figure 6). Participants started with an initial BML script and added modifications step by step following the assignments. After finishing the assignments participants could check their specification by comparing the resulting behaviour with that shown in a movie available on YouTube. The BML script had to be sent back to the evaluators and the participant had to fill in an online questionnaire.

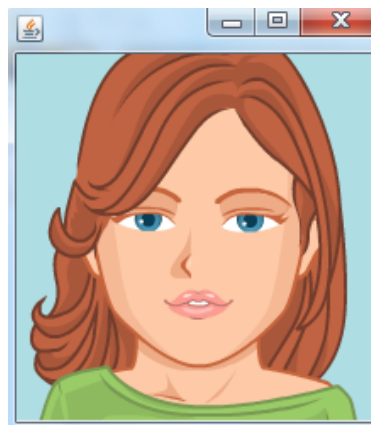


Figure 6. The ECA screen.

B. Materials

An initial BML script was prepared and given to the users. The participants had to add modifications to this script as described in the assignments. The assignments were selected in such a way that the participants had to discover all the functionalities of the PictureEngine available in the current version. The completed task would show the ECA introducing herself and let her do some physical exercises. Verbal and non-verbal behaviour should be synchronized together with different kind of gestures and gaze behaviour.

The online questionnaire that was filled in after the test is divided in several sections:

- 1) A section devoted to collecting some general background information about the users (gender, age, education, organisation in which the user works and his/her role in it, knowledge about ECAs and experience with designing/developing ECAs.
- 2) A section focussing on the process of making the assignments, in particular, how much time it took to complete all the assignments and how easy it was to complete the different parts of the assignment.
- 3) A section devoted to the usability of the user interface of the PictureEngine. In this section the users had to rate the clearness of different parts of the UI of the PictureEngine and were asked to provide suggestions for possible improvements.
- 4) A section focussing on the feedback provided by the PictureEngine, in which the users had to rate how clear they found the feedback messages provided by the PictureEngine.
- 5) A section where users were asked to give any further comments and suggestions.

C. Results

This section presents the results of the questionnaires and discusses the remarks returned by the participants.

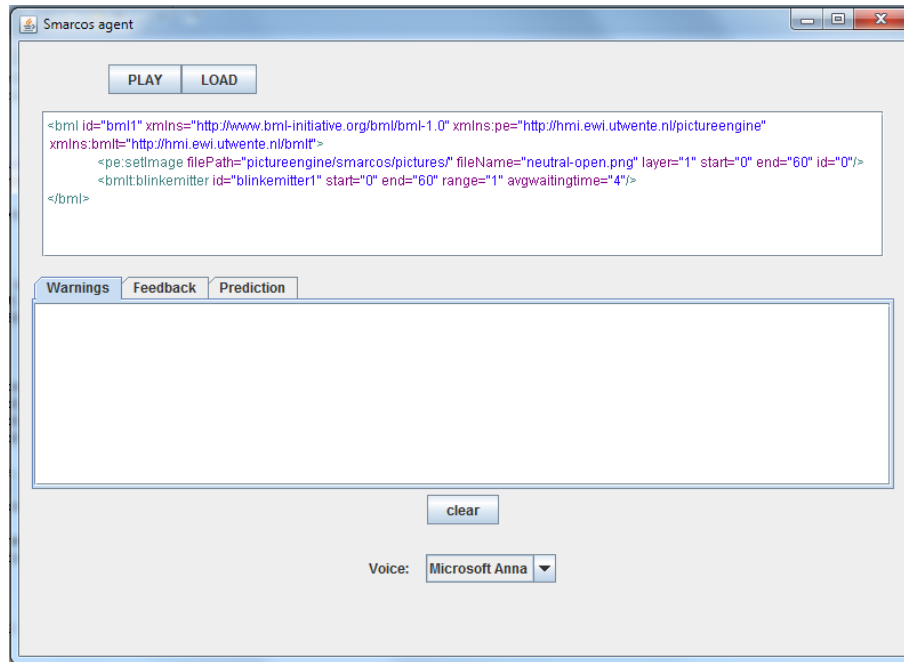


Figure 5. The BML editor screen

1) *Participants*: Twenty one persons participated in this user evaluation (5 females and 16 males), aged between 21 and 60, with an average age of 29. All participants were students or otherwise involved in research or education: from master students in Human Media Interaction, software engineers, to research or education in interface design and interactive systems, psychology and telemedicine. Twelve participants (5 female, 7 male) did not have any previous experience working with ECAs or specifying behaviour of embodied conversational agents and nine did have some experience with ECAs.

2) *Questionnaires*: Participants spend between 10 and 45 minutes to perform the whole user evaluation ($\mu = 25$, $\sigma = 9.4$) The participants were asked to fill in answers to the following questions on a 10-point Likert scale. The questions, mean values and standard deviations are presented below.

Participant had to rate statements about the easiness of adding gestures and gaze behaviour, the synchronisation of gestures and speech, and the assignment in general. Results of these questions can be found in table V-C2. Questions were answered on a 10 point Likert scale where 1 is “very hard” and 10 “very easy”.

From the figures in table V-C2 we conclude that overall the participants did not have many problems in completing the tasks.

3) *Remarks*: The participants were asked about what they found useful extensions of the capabilities to add sentences, gestures or gaze behaviours. Participants want to have an

Table I
SCORES OF COMPLEXITY QUESTIONS ON A 10 POINT LIKERT SCALE (1 IS VERY HARD, 10 IS VERY EASY)

Question	Mean	SD
How easy was it to add the gestures?	7.9	1.9
How easy was it to add the gaze behaviour?	8.1	2.0
How easy was it to synchronise gesture, gaze and speech?	7.5	2.5
How easy was it to finish the assignment?	8.0	1.5

auto-complete function and like to get some hints and tips while writing their BML script. To add gestures or gaze behaviour, participants like to select predefined gestures or gazes from a menu or list. The user interface of the editor tool should have bigger text field to enter and edit the BML script. Participants want to have a stop and pause button to stop or pause the execution of the BML script. The error messages from the Elckerlyc system should be less complex and should not show too much details and the participants want to see line number in the error message. Errors should be highlighted in the text field.

D. Conclusion

This small first user evaluation shows that BML together with the PictureEngine and the editor tool offers good possibilities to specify and test verbal and non-verbal behaviour of an ECA for mobile devices and can help developing more natural interactions with mobile user interfaces.

In general the participants were able to write and run their own BML script, even the participants that indicated not to have any experience with ECAs or did not have a technical background or programming skills.

From the comments and results of the questionnaires it became clear that the editor of the PictureEngine needs some improvements to make it more easy to create and run a BML script with respect to the design of the user interfaces, and feedback and error messages. The user evaluation with the PictureEngine made use of the general desktop user interface of the Elckerlyc platform. Changing the editor to a Eclipse plugin for better support and feedback while creating a BML script can be interesting improvement to the PictureEngine. Having a plug-in for Eclipse will also makes it possible to upload and run the BML script directly to a mobile Android device.

VI. APPLICATIONS

With the growing availability of online services and ubiquitous computing capabilities it becomes easier to develop systems that can present people information about their daily behaviours. This may help them to manage their lifestyle [21]. Sensor data and context information is available everywhere at any time. Some of these systems support people in their daily life by means of a human or digital coach. These systems can support users in coping with chronic diseases like COPD [22] and diabetics, but also to be more physical active [11][23]. Persuasive systems [24], and especially behaviour change support systems are information systems designed to form, alter or reinforce attitudes, behaviours or an act of complying without using deception, coercion or inducements [25].

The next sections present two different behaviour change support systems in which the PictureEngine was integrated in the user interface. Personalisation of the user interface by means of an ECA may affect the effectiveness of the behaviour change program and the user experience. Results from other studies indicate that the use of an ECA in a persuasive system has a positive effect on how the feedback is received by the user and on the results of the coaching program [26][27][28].

In the Smarcos application (see Section VI-A) the PictureEngine is integrated in the mobile Android devices that are part of the system. The C3PO system [22] described in Section VI-B integrated the PictureEngine on mobile Android devices and desktop PCs. Users were able to interact with the system on different systems and switch between the devices during the interaction.

A. Smarcos

In the EU Artemis project Smarcos we developed a personal digital health coach that supports users in attaining a healthy lifestyle by giving timely, context-aware feedback about daily activities through a range of interconnected

devices [29]. The two targeted user groups of the coaching system are office workers and diabetic type II patients. Office workers will receive feedback about their physical activity level, while diabetic type II patients also receive feedback about their medication intake. Physical activity is measured by a 3D accelerometer and medication intake is tracked by a smart pill dispenser. The pill dispenser is using the mobile network to connect to the internet. The system is context-aware and multi-device which means that the (digital) coach can support the users in various contexts and on different devices. GPS information is provided by the mobile phone of the user. The system sends feedback to the mobile phone of the user, the laptop or PC and their television.

Figure 7 shows the overall architecture of the coaching system. All information from the monitoring devices and manual input from the users are uploaded to the cloud and stored in a central knowledge base. The coaching engine contains coaching rules and continuously keeps track of all user data. When the coach receives a trigger it starts to evaluate the coaching rules. When one of the rules is evaluated with a positive result, it will select a suitable message from the coaching content database and send the message to the user through one of the available output devices. Output devices that are able to run a BML realiser, like the PictureEngine, are called BML enabled devices. BML enabled devices in the Smarcos coaching system are delineated in purple in figure 7. These devices are able to present feedback by an animated spoken interaction with an ECA. Feedback from the system can be presented as a text message, by means of a graph or by an ECA.

A first user evaluation with a basic version of the Smarcos personal digital health coach compared two alternatives for providing digital coaching to users of a physical activity promotion service. Participants in the study (N=15) received personalized feedback on their physical activity levels for a period of six weeks. Feedback was provided weekly either by e-mail or through an embodied conversational agent. The messages by the ECA were prerecorded video message. User's perception of the digital coaching was assessed by means of validated questionnaires after three weeks and at the end of the study. Results show significantly higher attractiveness, intelligence and perceived quality of coaching for the ECA coach.

B. C3PO

The Telemedicine group of Roessingh Research and Development (RRD) has over the past few years been working on a technology platform for supporting physical activity behaviour change in patients suffering from various chronic diseases, as well as for healthy individuals. This platform, called the *Continuous Care & Coaching PlatfOrm*, or C3PO, consists of a 3D-accelerometer based sensor, a Smart phone, back-end Server and connected Web portals (see Figure 8). The platform has been used successfully in trials with

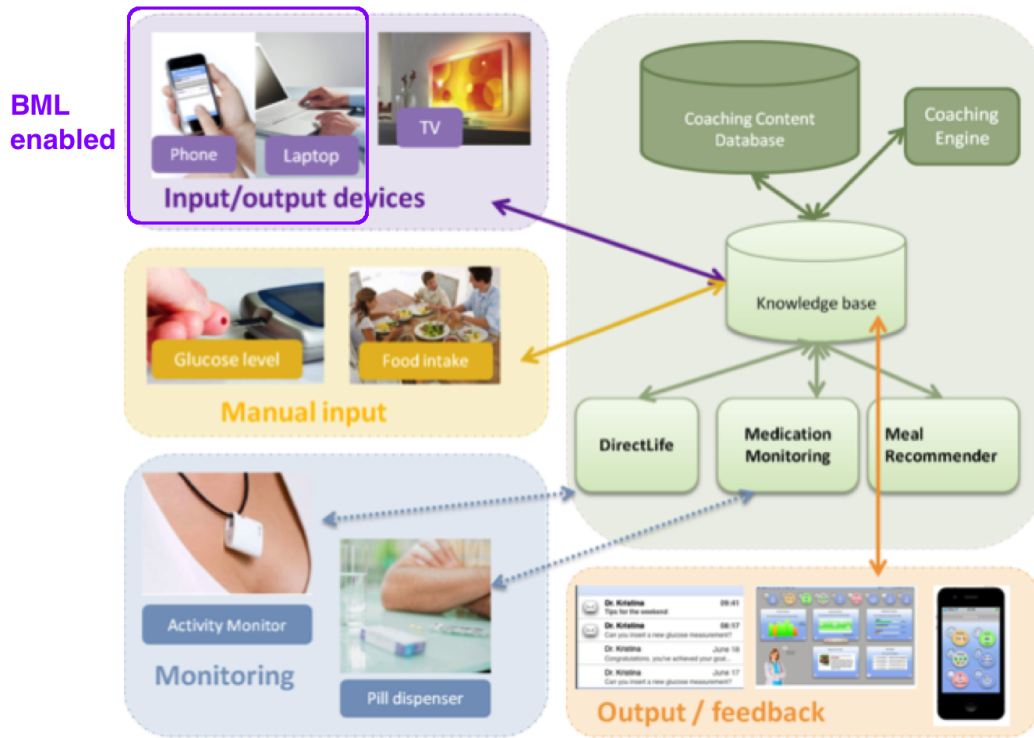


Figure 7. Overview of the Elckerlyc architecture. BML enabled output devices are marked in purple.

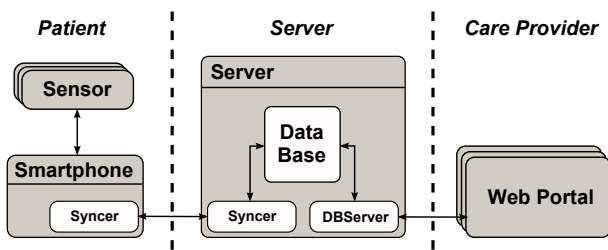


Figure 8. High level overview of the C3PO coaching platform [30].

patients suffering from Chronic Obstructive Pulmonary Disease (COPD), Chronic Low Back Pain (CLBP), Chronic Fatigue Syndrome (CFS), and Obesity [30], and is constantly under development to increase its effectiveness in real-time, tailored coaching. While earlier research focused on tailoring the delivery of motivational messages to the user in terms of timing and content (see, e.g., [22]), the visual representation of the feedback has been largely ignored. Due to the modular architecture of the Smart phone application, we were able to quickly integrate the PictureEngine in order to enable a more natural communication to the patient through the use of the ECA.

Two separate experiments were carried out to evaluate

the PictureEngine integrated with the C3PO platform. The first is a controlled experiment to evaluate user perception and experience with receiving feedback from the ECA compared to the regular text-based user interface. For the second experiment, a more complex system was developed, similar to the Smarcos system described in the previous section, where feedback was presented to the user on various interconnected devices, including two BML enabled devices (PC and Smart phone). For both experiments, the target user groups were healthy office workers.

The first user evaluation comparing the use of an ECA to the standard text feedback message interface included 14 participants, aged between 22 and 61 ($\mu = 37, \sigma = 13.3$) and consisting of 8 males and 6 females. Participants were randomly assigned to either the text-first condition, in which they received the standard text based interface in the first week, and the ECA in the second week, or the ECA-first condition. All participants finished the evaluation, with 1 participant not being able to complete full measurement days due to a faulty sensor. In both conditions the system generated a *motivational cue* message every hour, based on the user's current activity progress compared to a predefined reference activity pattern. When asked about the user's preference for either of the two conditions, only 3 users preferred the ECA, 10 users preferred the text-only condition, and 1

user had no preference. The single most important reason given by the users for preferring the text-only condition is *glanceability*. As the ECA pronounces the feedback message with real-time subtitling (letter for letter), it takes a much longer time to convey the entire message compared to the text-based interface, where users can read it immediately. When asked directly about the ECA, users responses were varied. Four participants found the ECA fun and enjoyable, and three participants said the ECA added personality to the system. On the negative side it was commented that the ECA does not add anything significant to the system, and that the ECA was not believable because it was not a real person. Two participants commented that the ECA did not show enough enthusiasm, while another thought the ECA was too enthusiastic to the point of it not being believable any more. With only 3 out of 14 users preferring the ECA over the simple text version it can be concluded that the result of the evaluation is not in favour of using ECA's in this application. However, besides the fact that obvious improvements can still be made (such as better graphics and voice output), an interesting observation is the dichotomy between user's opinion about the ECA as personification of the application. This dichotomy was also mentioned as a key outcome of a study on Embodied Agents in 1996 by Koda & Maes [31]. It seems that the perception of an ECA is highly personal. Comments regarding future improvements to the system from the users are in line with this, which include suggestions about making the ECA more relevant to the target audience, enabling users to choose between different visual/auditory styles of the ECA, and enabling the user to choose whether or not to use the ECA at all.

One of the advantages of ECAs as user interface is the personification aspect. At least for some users, the ECA gives a certain personality to an application. This personality aspect can be exploited in the area of multi-device user interfaces to overcome some of the challenges related to continuity. Multi-device user interfaces exist in many different forms and levels of complexity. Using Paternó & Santoro's [32] framework and terminology, our multi-device platform can be described as a system that supports *UI Migration*, automatic *Trigger Activation* and *Multi-modal* devices using a client/server architecture. For the second evaluation we have implemented a multi-device component to the C3PO platform. For the target population of office workers, the idea is that throughout a regular work day, the user communicates with various devices that each offer unique capabilities in terms of physical activity coaching. While performing desk work, the PC would be the most suitable device for delivering communication from the system to the user, if the user is getting a coffee, a public screen mounted next to the coffee machine can offer coaching through social influencing, and while the user is travelling, the Smart phone can take over the coaching role for ubiquitous availability. In order to accommodate the virtual coach migrating across

devices with the user, we have added a server component that manages so-called *user-requests*. Whenever a particular output device (PC, Smart phone, Public Screen) notices that the user is near it and available for communication, the device would send a user-request to the server. The server then decides which is the most suitable device for coaching, and notifies the devices of their ability to engage with the user. Also implemented for this evaluation is the ability to engage in short dialogues with the ECA on the PC or Smart phone. Users were presented with spoken questionnaires and were able to use speech input to answer the (multiple choice) questions. In order to accommodate user preferences regarding UI, the Smart phone supports either the ECA and a regular text-based interface for the questionnaires, between which the user can switch at any time. Six participants took part in the evaluation of this multi-device version of the coaching platform, 4 females and 2 males. The evaluation was small and focused on usability testing and a "thinking aloud" procedure. Users were observed while performing a set of tasks involving desktop computer work, going to the coffee machine, and walking around the office. Most of the findings from this early stage evaluation were related to the technical working of the system or missing functionalities, however from a usability perspective it again became clear that there are large differences between user preferences for the device selection, and user-interface selection.

From our experiences with both these applications, it became clear that the use of the PictureEngine can be a useful tool for tailoring the user interface. However, due to the large differences in users, the option of switching to a more classic interface (either automatically, or through user selection) should always be supported. As with other forms of tailoring, such as automatically adapting the coaching style (formal vs. informal), it remains an open problem how to automatically match the right user interface representation to the right user.

VII. CONCLUSION AND FUTURE WORK

To take the full advantage of the known benefits of personification of the user interface of service systems a mobile platform that is able to present embodied conversational agents in mobile applications is presented. The platform makes use of the Elckerlyc system. Because it is too heavy to render realistic 3D virtual humans on mobile devices a light-weight PictureEngine was developed. The PictureEngine makes it possible to use the Elckerlyc system on the Android platform and to generate real-time animations of embodied conversational agents.

The PictureEngine is applied as user interfaces in three coaching applications. Initial results of short term user evaluations showed a large variety in the response of users to the ECA, in line with earlier research in this area by Koda & Maes [31]. As for some users the ECA was perceived very positively (adding personality to the system and increasing

the feeling of consistency in the multi-device application), we believe that the use of ECA's should be offered as optional component in such coaching applications. Also, additional work is needed in allowing the user to make personal choices regarding the appearance of the ECA, both visually and in terms of voice and character.

When such modifications are in place, long term user evaluations with these coaching platforms, including the mobile embodied coach, are planned to investigate the effects of personalised coaching feedback on user experience, quality of coaching and effectiveness of the coaching program.

The results of user evaluation showed that BML together with the PictureEngine and the editor tool offers good possibilities to specify and test verbal and non-verbal behaviour of an ECA for all participants. All the participants, including the non expert users without any programming skills were able to complete the assignment.

Although it is shown that the PictureEngine can run on mobile Android devices it would be worth exploring options for using a different TTS system in the future. This would allow to regain the speech-related functionality that is currently unavailable on Android, such as synchronization within utterances and viseme-based lip-sync. A next step in the development of the PictureEngine is looking for techniques to allow small movements by the ECA, such as nodding and shaking of the head. These movements will make it possible for the ECA to show those communicative behaviours that have shown to be effective for turn-taking and attention signalling in real-time interactions.

ACKNOWLEDGMENT

This work was funded by the European Commission, within the framework of the ARTEMIS JU SP8 SMARCOS project 100249 - (www.smarcos-project.eu, last access date 28-06-2013).

The PictureEngine application for Android is released under GPL3 license. The PictureEngine is an Android implementation of the Asap BML realizer which is already available through <http://asap-project.ewi.utwente.nl> (last access date 28-06-2013).

REFERENCES

- [1] R. Klaassen, J. Hendrix, D. Reidsma, and H. J. A. op den Akker, "Elckerlyc goes mobile: enabling technology for ECAs in mobile applications," in *The Sixth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, Barcelona, Spain*. XPS (Xpert Publishing Services), September 2012, pp. 41–47.
- [2] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone, "Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents," in *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, ser. SIGGRAPH '94. New York, NY, USA: ACM, 1994, pp. 413–420.
- [3] T. Bickmore, D. Mauer, F. Crespo, and T. Brown, "Persuasion, task interruption and health regimen adherence," in *Proceedings of the 2nd international conference on Persuasive technology*, ser. PERSUASIVE'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 1–11.
- [4] M. Argyle, *Bodily Communication*, MethuenEditors, Ed. Methuen, 1988, vol. 2nd.
- [5] D. Traum, S. Marsella, J. Gratch, J. Lee, and A. Hartholt, "Multi-party, multi-issue, multi-strategy negotiation for multimodal virtual agents," in *Intelligent Virtual Agents*, ser. Lecture Notes in Computer Science, H. Prendinger, J. Lester, and M. Ishizuka, Eds. Springer Berlin / Heidelberg, 2008, vol. 5208, pp. 117–130.
- [6] H. van Welbergen, D. Reidsma, Z. M. Ruttkay, and J. Zwiers, "Elckerlyc - a BML realizer for continuous, multimodal interaction with a virtual human," *Journal on Multimodal User Interfaces*, vol. 3, no. 4, pp. 271–284, August 2010.
- [7] C. E. Garber, B. Blissmer, M. R. Deschenes, B. A. Franklin, M. J. Lamonte, I.-M. Lee, D. C. Nieman, and D. P. Swain, "American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise." *Medicine & Science in Sports & Exercise*, vol. 43, no. 7, pp. 1334–1359, 2011.
- [8] F. Buttussi and L. Chittaro, "Mopet: A context-aware and user-adaptive wearable system for fitness training," *Artificial Intelligence in Medicine*, vol. 42, no. 2, pp. 153 – 163, 2008.
- [9] M. Turunen, J. Hakulinen, O. Stahl, B. Gambäck, P. Hansen, M. C. R. Gancedo, R. S. de la Cmara, C. Smith, D. Charlton, and M. Cavazza, "Multimodal and mobile conversational health and fitness companions," *Computer Speech & Language*, vol. 25, no. 2, pp. 192 – 209, 2011.
- [10] M. W. Kadous and C. Sammut, "Inca: A mobile conversational agent," in *PRICAI 2004: Trends in Artificial Intelligence*, ser. Lecture Notes in Computer Science, C. Zhang, H. W. Guesgen, and W.-K. Yeap, Eds. Springer Berlin / Heidelberg, 2004, vol. 3157, pp. 644–653.
- [11] O. Ståhl, B. Gambäck, M. Turunen, and J. Hakulinen, "A mobile health and fitness companion demonstrator," in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics: Demonstrations Session*, ser. EACL '09. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, pp. 65–68.
- [12] (2013) Mindmakers.org. [Online]. Available: <http://www.mindmakers.org> (last access date 28-06-2013)
- [13] E. Bevacqua, K. Prepin, E. de Sevin, R. Niewiadomski, and C. Pelachaud, "Reactive behaviors in SAIBA architecture," in *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, S. Decker, Sichman and Castelfranchi, Eds., May 2009.

- [14] H. Vilhjalmsson, N. Cantelmo, J. Cassell, N. E. Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud, Z. Ruttkay, K. R. Thórisson, H. V. Welbergen, and R. J. V. D. Werf, "The behaviour markup language: recent developments and challenges," in *Proceedings of the 7th International Conference on Intelligent Virtual Agents*, ser. Electronic Notes in Artificial Intelligence, vol. 4722. Springer, 2007, pp. 99–111.
- [15] A. Heloir and M. Kipp, "Real-time animation of interactive agents: specification and realization," *Applied Artificial Intelligence*, vol. 24(6), pp. 510–529, 2010.
- [16] E. Bevacqua, M. Mancini, R. Niewiadomski, and C. Pelachaud, "An expressive ECA showing complex emotions," in *AISB'07 Annual convention, workshop "Language, Speech and Gesture for Expressive Characters"*, April 2007, pp. 208–216.
- [17] D. Reidsma, I. de Kok, D. Neiberg, S. Pammi, B. van Straalen, K. Truong, and H. van Welbergen, "Continuous interaction with a virtual human," *Journal on Multimodal User Interfaces*, vol. 4, no. 2, pp. 97–118, 2011.
- [18] D. Reidsma and H. v. Welbergen, "Elckerlyc in practice on the integration of a BML realizer in real applications," in *Proc. of Intetain 2011*, 2011.
- [19] D. Heylen, M. Theune, H. J. A. op den Akker, and A. Nijholt, "Social agents: the first generations," in *ACII 2009: Affective Computing & Intelligent Interaction*, M. Pantic and A. Vinciarello, Eds. Los Alamitos: IEEE Computer Society Press, September 2009, pp. 114–120.
- [20] E. Bevacqua, M. Mancini, R. Niewiadomski, and C. Pelachaud, "An expressive eca showing complex emotions," in *AISB'07, Artificial and Ambient Intelligence*, Newcastle, UK., 2007.
- [21] M. Kasza, V. Szücs, A. Vég, and T. Török, "Passive vs. active measurement: The role of smart sensors," in *UBI-COMM 2011, The Fifth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, 2011, pp. 333 – 337.
- [22] H. op den Akker, V. Jones, and H. Hermens, "Predicting feedback compliance in a teletreatment application," in *Proceedings of ISABEL 2010: the 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies*, 2010.
- [23] G. Geleijnse, A. van Halteren, and J. Diekhoff, "Towards a mobile application to create sedentary awareness," in *In Proceedings of the 2nd Int. Workshop on Persuasion, Influence, Nudge Coercion Through Mobile Devices (PINC2011)*, vol. 722, 2011, pp. 90–111.
- [24] B. Fogg, *Persuasive Technology. Using computers to change what we think and do.* Morgan Kaufmann, 2003.
- [25] H. Oinas-Kukkonen, "Behavior change support systems: A research model and agenda," in *Persuasive Technology*, ser. Lecture Notes in Computer Science, T. Ploug, P. Hasle, and H. Oinas-Kukkonen, Eds. Springer Berlin / Heidelberg, 2010, vol. 6137, pp. 4–14.
- [26] O. A. B. Henkemans, P. van der Boog, J. Lindenberg, C. van der Mast, M. Neerinx, and B. J. H. M. Zwetsloot-Schonk, "An online lifestyle diary with a persuasive computer assistant providing feedback on self-management," *Technology & Health Care*, vol. 17, p. 253–257, 2009.
- [27] D. Schulman and T. W. Bickmore, "Persuading users through counseling dialogue with a conversational agent," in *Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology*. New York, NY, USA: ACM, 2009, pp. 1–8.
- [28] D. C. Berry, L. T. Butler, and F. de Rosis, "Evaluating a realistic agent in an advice-giving task," *Int. J. Hum.-Comput. Stud.*, vol. 63, no. 3, pp. 304–327, 2005.
- [29] R. op den Akker, R. Klaassen, T. Lavrysen, G. Geleijnse, A. van Halteren, H. Schwietert, and M. van der Hout, "A personal context-aware multi-device coaching service that supports a healthy lifestyle," in *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, ser. BCS-HCI '11. Swinton, UK, UK: British Computer Society, 2011, pp. 443–448.
- [30] H. op den Akker, M. Tabak, M. Marin-Perianu, R. M. Huis In't Veld, V. M. Jones, D. Hofs, T. M. Tonis, B. W. van Schooten, M. M. Vollenbroek-Hutten, and H. J. Hermens, "Development and Evaluation of a Sensor-Based System for Remote Monitoring and Treatment of Chronic Diseases - The Continuous Care & Coaching Platform," in *Proceedings of the Sixth International Symposium on e-Health Services and Technologies (EHST 2012)*, Geneva, Switzerland, 2012, pp. 19–27.
- [31] T. Koda and P. Maes, "Agents with Faces : The Effect of Personification," in *Proceedings of the 5th IEEE International Workshop on Robot and Human Communication*, 1996, pp. 189–194.
- [32] F. Paternò and C. Santoro, "A logical framework for multi-device user interfaces," *Proceedings of the 4th ACM SIGCHI symposium on Engineering interactive computing systems - EICS '12*, p. 45, 2012.