E-Composer: Enabling the Composition of Mobile Assistants

Ilhan Aslan*, Dyuti Menon*, Robert Brauer*, Kristin Albert* and Christian Maugg*
Fraunhofer ESK, Germany*
name.lastname@esk.fraunhofer.de*

ABSTRACT

ELEPHANT (ELEments for Pervasive and Handheld AssistaNTs) is a system that aims to integrate a broad range of users (e.g. designers, domain experts and end users) with different backgrounds in the process of developing personal mobile assistants. In this paper we present a user study that we have conducted for two reasons: First, to screen characteristics of modeling mobile assistants by non-experts of mobile software development; and second, to test a first prototype of the ELEPHANT system's graphical modeling tool (E-Composer).

1. INTRODUCTION

Today, the use of mobile phones is very wide spread. In addition, the capabilities of mobile technology as also the underlying infrastructure are increasing on a regular basis. This development qualifies mobiles phones as digital companions in everyday life. However, when it comes to modeling the interaction for a broad spectrum of target users, target domains and context of use, the modeling process becomes very cumbersome. On the one hand, designing interaction and user interfaces is a profession in itself and most software engineers do not have the required skills to build user centered, attractive and usable interactions without being guided or having a framework set for them. On the other hand, general modeling languages (e.g. UML based) that are being used by software engineers are either too low level or foreign to most designers and domain experts. The ELEPHANT (ELEments for Pervasive and Handheld AssistaNTs) system aims to integrate non-software engineers (e.g. designers, domain experts and end users) in the process of developing personal mobile assistants. The ELEPHANT system's modeling tool that we refer to as the E-Composer allows a high level of modeling based on components [1]. One of the reasons why users access services while mobile is basically because they need assistance to complete an activity (e.g. shopping, dining, driving or route finding) or to proceed with an activity in the real world. Although today's mobile phones have advanced interfaces and can handle most websites that have been originally designed for the desktop environment, single services that focus on content and functionality are not sufficient in assisting mobile users during their specific activities. Especially, if users are involved in real world activities in which they are pressed for time, the assistance provided through the capabilities of the mobile phones has to be highly personalized and centered to the user's activity. The requirements on personalization and adaptation to user activities are very high. To fulfill these requirements, domain experts and end users have to participate in the design process. Therefore, the ELEPHANT system provides a browser based tool support for the participative design of mobile assistants. The E-Composer is the front-end of the ELEPHANT system that allows users to graphically compose mobile assistants based on components. The graphical presentation of a mobile assistant modeled with the E-Composer has a tree-like structure (see figure 1). The backend of the ELEPHANT system manages these components. Components can be accessed and tagged with information by all users. Users can search for components and they can set up a components library. In [1] we described the component based development of mobile assistants in more detail.

In order to derive essential feedback regarding the ELEPHANT's composer tool, its reception by users and its functionalities, we describe in this paper usability tests that we conducted to measure user satisfaction from working with the tool and the overall performance of the tool. A small test scenario was setup, where users were given the task of modeling a mobile assistant using the ELEPHANT composer. Based on the user reactions and suggestions during and after the tests, conclusions were drawn regarding the performance and efficacy of the composer and how it may be improved. In this paper we present a description about the usability tests, the set-up and the data, what we intend to deduce from these usability tests and what methods we used to evaluate the data.

2. User Study

The usability tests were conducted with 11 participants in the age group of 22 – 28 years. They came with different backgrounds in the areas of computer expertise, authoring systems and system modeling skills. The tests were conducted individually and in an undisturbed setting with the test subject being initially instructed as to the nature and goal of the test. The test subjects were advised to complete the test within 1 hour and to keep in mind that this test was composed of 2 separate tasks. Once the test subjects were given all the instructions and provided with all the material to proceed with the test, the members of our team left the premises The goal of the tests was for the participants to create a mobile assistant, which would assist a friend who would shortly be travelling to the city of Barcelona. This mobile assistant would aid the visitor with the Spanish language by helping them with the translations of common phrases (to buy tickets, order food etc.), be a guide for sightseeing in the city of Barcelona (by providing background information on the interesting places to see) and provide additional information such as suggestions about interesting places to eat or things to do in Barcelona. Keeping the generation of a Barcelona mobile assistant as the common goal, two tasks were designed to differentiate between a known and an unknown framework. The first task was to design a paper based Barcelona mobile assistant (see figure 2). The second task was to do the same, i.e. design a Barcelona mobile assistant, with the

Pre-proceedings of the 5th International Workshop on Model Driven Development of Advanced User Interfaces (MDDAUI 2010): Bridging between User Experience and UI Engineering, organized at the 28th ACM Conference on Human Factors in Computing Systems (CHI 2010), Atlanta, Georgia, USA, April 10, 2010.

help of the ELEPHANT composer (see figure 1). For both the tasks, the test subjects were provided with a list of content they had at their disposal to create this assistant. The content included text data, images, video clips and audio files, all connected to Barcelona and the Spanish language.

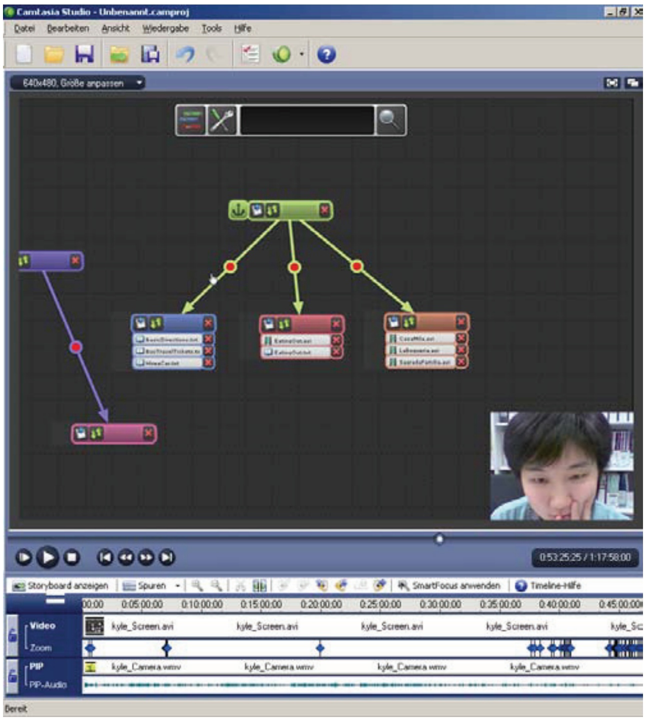

Figure 1: Screenshot of one of the subject's audio and video data

Figure 2: Photo of a result of one of the subject's paper based model of a mobile assistant

Our aim in conducting these tests was to measure the system performance, user satisfaction and the emotional response (in terms of stress and cognitive load on the participant) due to using the tool. System performance: Evaluating the operation and efficiency of the tool is a key step in its development. Identifying areas that require more attention or areas that we can build up on help enrich the authoring tool and provide a solid basis to create an advanced product. User satisfaction: Based on actual user experience, this metric is a powerful indicator of how the product might be received and how quickly it might be adopted by users. The test subjects rate and rank different features and functionalities of the tool and we as developers are able to interpret this and change and improve the authoring tool accordingly. Indication of stress and Cognitive Load: The term cognitive load (CL) may be described as the amount of effort that accompanies learning, thinking and reasoning [9] and hence has a bearing on the overall evaluation of the tool.

System performance and user satisfaction: In our usability tests, both these metrics were evaluated from user feedback in the form of questionnaires, user comments and user reactions. Real-time user reactions were also recorded by capturing the screen activity, recording any comments made by the test subjects while doing the tests and by using a webcam to record the activity of the test subjects (see figure 1). Stress and Cognitive Load: As discussed earlier, both stress and cognitive load introduce physiological changes in body, they can be identified using biosensors that monitor and record certain bio-signals. In our usability tests, we monitored the heart rate, skin conductivity and skin temperature of our test subjects.

3. Data Collection

Two questionnaires were administered to the users. The first was used to understand the background of the user and his experience with any of the authoring tools available in the market. This was answered by the test subject before beginning the usability test. The second questionnaire addressing issues related to the ELEPHANT Composer was answered by the test participants after the completion of both the tasks. This one was largely based on the USE Questionnaire for User Interface satisfaction, designed by Arnold Lund [6]. This particular questionnaire evaluates four key factors, Usefulness, Ease of Use, Ease of Learning and Satisfaction, through a series of questions, which are answered by rating (from 1 to 7) between a strongly positive reaction (scored as 7) to a strongly negative one (scored as 1). Test subjects were also given the freedom to express their suggestions and ideas. The test subjects were asked to think aloud and a continuous audio and video recording was made, whereby we could register their thoughts and reactions during the course of the task. In order to correlate these audio comments with the task being performed, the activity on the screen was also captured with the help of Camtasia Studio 5, Screen Recording Software. Using Camtasia we were also able to record the video feed from a webcam that was monitoring the test subject (see figure 1). All these 3 inputs were recorded to be part of the usability test analysis.

In our study, we intended to measure changes in 3 physiological variables, namely heart rate (indicator of stress), skin conductivity (or electrodermal activity [3] - an indicator of CL) and skin temperature (indicator of stress). To carry out these measurements we used two biosensors, the Alive Technologies Heart Monitor and the SenseWear BMS from Body Media. We monitored the bio-signals of the test subjects over both the tasks, allowing us to compare levels of parameters such as CL or stress between the paper-based and tool-based task.

4. Data Interpretation

An initial questionnaire was answered by the test subjects at the start of the test to ascertain the level of computer knowledge and experience with authoring tools and system modeling. Since the test subjects' profession ranged from computer scientists to economists and electrical engineers, we have encountered different levels of both computer knowledge and designing and modeling experience. However, all participants estimated themselves as being capable of operating personal computers, while the self-assessment regarding the experience with software modeling and authoring tools varied quite a lot between the test subjects. We were expecting to see reduced cognitive load for participants with a high level of knowledge regarding software modeling and authoring tools. The second questionnaire (based on the USE Questionnaire for User Interface) was administered after the completion of both the tasks. The second questionnaire was evaluated based on the guidelines as set by the author, and gave us an insight into the levels of user satisfaction and ease of use of the composer. The audio and video recording was evaluated in conjunction with the task that was being performed at that time. The comments made were interpreted along with the activity occurring on the screen and the webcam feed recorded within that time frame, to see what it was about our tool that caused them to have a problem and to see if they had any suggestions to change and improve the tool. As our aim was to analyze the cognitive load (the evaluation of stress is a part of our future work) on the test subjects and depending on the findings, find ways to improve the tool, making it easier to use. To this effect, we analyzed the Galvanic Skin Response (GSR) values tracked by the SenseWear BMS biosensor. We performed a simple statistical analysis, calculating the mean over the entire test duration and over each of the tasks separately. Any task which requires learning, thinking and/or reasoning, puts a certain amount of load on the working memory, known as Cognitive Load (CL) [8]. There are 3 types of CLs associated with learning a task. The intrinsic CL is the inherent difficulty and complexity associated with a task. The extraneous CL is produced based on the manner in which the instruction or information is presented to the student and must be minimized for optimum learning. Finally, the germane CL also originates from the manner of instruction, but contributes towards the learning process [8]. As the number of issues that can be simultaneously handled by the working memory is limited, the Cognitive Load Theory (CLT) provides a basis for designing optimum instructional interfaces which reduces the extraneous CL thereby ensuring more effective learning [7]. A lot of work has been done on using CL to reduce the difficulties associated with learning computer programming which is a highly interactive task. More interaction increases the CL on the working memory as multiple activates and skills are being called upon simultaneously [10]. For tasks rich in interactivity, it is particularly important to reduce the extraneous CL [8]. As in [9] we use the GSR data obtained from our biosensors in order to analyze the effect of CL on our participants, as there is a directly proportional correlation between the GSR values and CL (an increase in CL results in an increase in the GSR [9] and vice versa). Out of the 11 participants, 9 were chosen for the analysis of biosensor data (the data for the other 2 participants was not collected as planned due to problems with improper skin contact).

For the analysis, the entire duration of the test was split up into 3 parts (see figure 3), namely:

- Listening to instructions: where the participants received the initial instructions, including a brief description of the test and the goals
- Paper Based task: where the participant carried out the paper-based task (not time limited) to design a mobile travel assistant on paper
- Computer Based task: where the participants used the ELEPHANT composer to create the same travel assistant

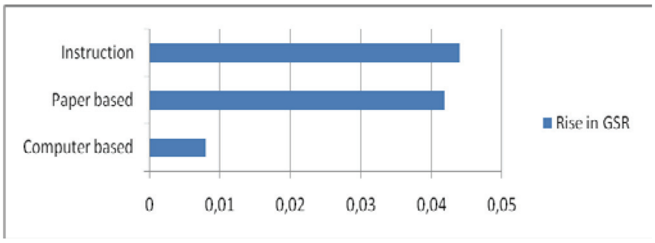

Figure 3: Rise of GSR in µS for participant Banner, opposed for each of the three individual parts (instruction, paper based and computer based)

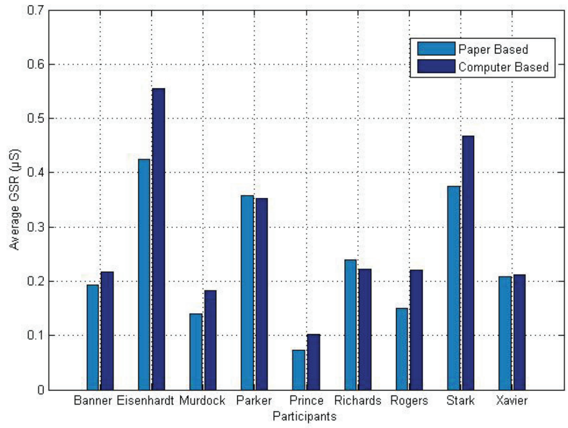

Figure 4: Average GSR for paper based and computer based tasks for each participant

The SenseWear BMS from Body Media provided us with a moving average of GSR for every minute over the entire duration of the test. As each participant spent variable amounts of time on each of the tasks, we calculated the mean GSR for each of the above time intervals for each participant, which allowed us to compare these values.

$$avgGSR_{task}(i) = \frac{\sum GSR_{task}(i)}{t_{task}(i)} \quad (1)$$

where t_{task} is the duration of each task, i represents the participant and GSR_{task} represents the recorded moving average GSR values for the task being undertaken (listening to the instructions, working on the paper-based, or using the composer). The mean GSR values of the paper-based and computer-based tasks for each of the participants were then compared. Based on these metrics, we present our results in the next section.

5. Conclusion and Future Work

Using the composer people felt comfortable with the system and recommended the quiet simple use of its interface. User-friendliness and the ease of learning were also appreciated by most of the participants. All participants succeeded in searching for resources and arranging them to an expected final structure with marginal variations based on the respective level of creativity and effort put into the application. A limited scale of ELEPHANT elements (E-elements) provided from the system within the testing scenario delimitated freedom of choice. Participants felt restricted of the predetermined set of E-elements. They desired a drilldown of basic E-elements with the possibility to vary these items according to their goals.

Once the mean GSR for each participant for each of the tasks was calculated, we performed the following comparisons to deduce the CL generated in our test subjects, due to using our tool. The average GSR for the 3 tasks of the usability tests were as follows: listening to instructions 0.18 µS, paperbased 0.24 µS and computer based 0.28 µS. As expected, there was an increase in the average GSR for the computer based task, indicating an increase in the CL. This clearly supports the theory that moving from a known environment (paper based) to an unknown environment (the ELEPHANT Composer) which involves the usage of a new computer tool causes a rise in the cognitive load on the memory. The next step was to examine the average GSR for each of the participants individually. As we are specifically interested in the paper based and computer based tasks, figure 4 plots the average GSR calculated for each participant in these 2 tasks. In order to see the significance of the change (increase or decrease), we also calculated the change in the average GSR in the computer based task with respect to that of the paper based task and expressed it as a percentage.

$$Change\ \% = \frac{avgGSR_{computer}(i) - avgGSR_{paper}(i)}{avgGSR_{paper}(i)} \times 100 \qquad (2)$$

where *i* is represents each participant. While the general trend is to have an increase in the GSR (and hence an increase in CL), we observed that for 2 participants (Richards and Parker) there was a decrease in the GSR recorded during the computer based test. Comparing the GSR results with those of the questionnaires, we saw that Richards and Parker, both hailing from background of IT and with extensive computer expertise and experience in using authoring systems found our tool easy to use and were able to learn the use of it quickly. This was expected, as we have already noticed the test subjects' varying knowledge level in software modeling and authoring, as pointed out above. The CL that was exerted on their working memories reduced during the computer based task.

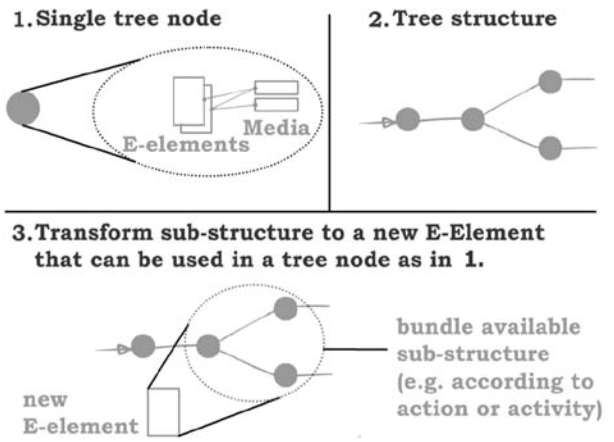

Figure 5: Bundling of substructures in tree nodes

In [1] we defined an ELEPHANT element (E-element) as a component with application logic. E-elements could only be developed by software engineers or designers with scripting abilities. We are planning to allow that new E-elements can also be composed with the E-Composer (see figure 5). With this improvement, the modeling based on components becomes more flexible but still keeps the high level. Because of the flexibility we gain, we also approach our long term goal of supporting activity-based design. Activities are dynamic and hierarchical structures. In activity theory, the objective of an activity can be realized through different sets of actions [5], different people might need different actions for the same activity and hence different ways to model the assistance for the same activity. Same actions can contribute to different activities, and may also have different meanings for the people undertaking them [4].

6. ACKNOWLEDGMENTS

This work was funded in part by the Bavarian Ministry of Economic Affairs, Infrastructure, Transport and Technology within the project „Dynamische Plattformen für Verteilte Systeme“.

7. REFERENCES

[1] I. Aslan and D. Menon. Component-based development of mobile assistants with the elephant system. In Proceedings of Mobility 2009, Nice, France, September 2-4, 2009.

[2] Elliot, S. N. et al., Cognitive load theory and universal design principles: Applications to test item development, Vanderbilt University, NASP Session, 2009

[3] Haag, A., Goronzy, S., Schaich, P., and Williams, J. Emotion recognition using bio-sensors: First steps towards an automatic system.2004, pp. 36-48.

[4] K. Kuutti. Activity theory as a potential framework for human-computer interaction research. In Context and Consciousness: Activity Theory and Human-computer Interaction, pages 17–44, 1996.

[5] A. Leont'ev. Activity, Consciousness, and Personality. Prentice Hall, New Jersey, 1978.

[6] Lund, A. Measuring usability with the use questionnaire, stc usability sig newsletter, 8:2.

[7] Oviatt, S., Human-Centred Design meetns Cognitive Load Theory: Designing interfaces that help people think, Proceedings of the 14th annual ACM international conference on Multimedia, 2006, pp. 871-880

[8] Richard E. M., The Cambridge handbook of multimedia learning, Cambridge University Press, 2005

[9] Shi, Y., Ruiz, N., Taib, R., Choi, E., and Chen, F. Galvanic skinresponse (gsr) as an index of cognitive load. In CHI '07: CHI '07 extended abstracts on Human factors in computing systems (New York, NY, USA, 2007), ACM, pp. 2651-2656.

[10] Yousoof, M., Sapiyan, M., and Kamaluddin, K., Reducing cognitive load in learning computer programming, World Academy of Science and Technology, Volume 12, 2006, ISSN 1307-6