

# Talking Faces in Lab and Field Trials

## A View on Evaluation Settings and User Involvement Results of Avatar Based User Interaction Techniques in Three Ambient Assisted Living Projects

Miroslav Sili<sup>1</sup>(✉), Jan Bobeth<sup>2</sup>, Emanuel Sandner<sup>1</sup>, Sten Hanke<sup>1</sup>,  
Stephanie Schwarz<sup>2</sup>, and Christopher Mayer<sup>1</sup>

<sup>1</sup> Health and Environment Department, Biomedical Systems, AIT Austrian  
Institute of Technology GmbH, Vienna, Austria

miroslav.sili@ait.ac.at

<sup>2</sup> Innovation Systems, Technology Experience, AIT Austrian Institute  
of Technology GmbH, Vienna, Austria

**Abstract.** In recent years, there has been an increasing interest in Ambient Assisted Living technology to support older adults. Research and industry are working jointly on reliable and suitable solutions to help older adults to remain healthy and safe while living independently. Appropriate interaction methods play an important role for the acceptance of such supporting systems. Today, solutions mainly rely on common and well-evaluated interaction techniques such as TV remotes or touch screens to enhance the usability. Projects presented in this work are based on the same interaction techniques, but additionally enrich the interaction experience with a real-time, empathic virtual assistance avatar. In this paper, we present evaluation settings and user involvement results acquired from three different Ambient Assisted Living projects focusing on avatar-based user interaction. Our results show that avatar-based interaction in the Ambient Assisted Living context is very well applicable, especially when combined with speech recognition.

**Keywords:** Avatar · User interaction · Ambient assisted living · Multimodality

## 1 Introduction

This work summarizes evaluation results of prototypes developed in the projects AALuis<sup>1</sup>, ibi<sup>2</sup> and Miraculous-Life<sup>3</sup> with special focus on avatar aspects. We present lab and field trial outcomes from the accomplished projects AALuis and ibi as well as interim lab trial results from the still ongoing project Miraculous-Life. All three

---

<sup>1</sup> <http://www.aaluis.eu>.

<sup>2</sup> <http://www.ibi.or.at>.

<sup>3</sup> <http://www.miraculous-life.eu>.

projects use avatar-based output in order to support the end user during the interaction with the system. Prototypes developed within the ibi and Miraculous-Life projects also include speech recognition, thus allow a smooth user-system interaction.

The prototypical outcome of AALuis is an open middleware layer that can be used to connect various AAL services to a variety of devices with the help of automatically generated user interfaces [1–4].

The overall goal of the ibi prototype is to provide a tailored communication platform which is easy to understand and which can easily be used by three target groups, namely older adults, their relatives and professional caregivers.

The aim of the Miraculous-Life project is to design, develop and evaluate a Virtual Support Partner (VSP) assisting older adults throughout their daily activities and safety needs [5].

## 2 Methodology

To guarantee high usability and user acceptance we followed a user-centered approach in all three projects [6]. User involvement took place in the concept phase as well as in the two prototype evaluation phases. This section gives an overview about the evaluation settings, the involved user groups and the evaluation methods.

### 2.1 Phase One: Evaluation Settings During the Concept Phase

Before the first phase, user wishes, needs and requirements were collected with the help of cultural probing. Results were discussed in several workshops to identify the needed functional requirements for the planned prototypes. Afterwards, we involved different user groups into the design process by discussing scenarios and conducting usability studies (e.g., thinking aloud [7, 8]) with early mock-ups. Additionally, service developers and providers, user interface designers as well as experts from care organization helped to classify requirements and to identify usability problems in the early development stage. Table 1 shows a comparison of the evaluation settings from the projects.

### 2.2 Phase Two: Evaluation Settings for the Lab Trials

The second phase was conducted in a similar way as the first phase. Different user groups evaluated the first running prototypes which already covered most of the previously defined user requirements. A limitation on functionality and offered services during the lab trials allowed us to rather focus on general system characteristics like usability, performance, complexity and efficiency. Table 2 summarizes the evaluation settings during the lab trials.

### 2.3 Phase Three: Evaluation Settings for the Field Trials

In this phase, final versions of the prototypes were tested during a period of 2 to 6 weeks. Since Miraculous-Life is still an ongoing project, we evaluated only two of the

**Table 1.** Comparison of evaluation settings during the concept phase

	AALuis	ibi	Miraculous-Life
Period	Dec. 2012 – Jan. 2013	Nov. 2012 – Feb. 2013	Mar. 2014 – Sep. 2014
Participants	12	26	41
Sex F/M	7/5	16/10	25/16
Average age	74.5	70 <sup>a</sup>	79.7; 50.2 <sup>b</sup> ; 32.5 <sup>c</sup>
User groups	Older adults, younger & active seniors	Older adults, informal caregivers	Older adults, formal and informal caregivers, experts
Methods	Structured interview, task-based empirical investigation, System Usability Scale SUS[9]	Structured interview, task-based empirical investigation, questionnaire	Structured interview, questionnaire, expert walkthrough [10], heuristic analysis [11]
Devices	Smartphone, PC, tablet, TV	Smartphone, tablet, TV	PC, tablet, kinect

<sup>a</sup> Age calculated based on the average value of the estimated participants age

<sup>b</sup> Formal and informal care givers

<sup>c</sup> Experts

**Table 2.** Comparison of the evaluation settings for the lab trials

	AALuis	ibi	Miraculous-Life
Period	Apr. 2013	Feb. 2014	Sep. 2014
Participants	9	17 <sup>a</sup>	18
Sex F/M	3/6	4/3 <sup>b</sup>	13/5
Average age	71.2	72.1 <sup>b</sup>	81.71; 45.6 <sup>b</sup>
User groups	Older adults, younger & active seniors	Older adults, informal and formal caregivers	Active older adults, formal caregivers
Methods	Task-based empirical investigation, SUS	Task-based empirical investigation, post-interviews, questionnaires	Task-based empirical investigation, think aloud method, questionnaire (SUS)
Devices	Smartphone, PC, tablet, TV	Smartphone, tablet, TV	PC, tablet, Kinect

<sup>a</sup> Sex and age of users in the user group older adults have not been elaborated

<sup>b</sup> Formal and informal care givers

**Table 3.** Comparison of the evaluation settings for the field trials

	AALuis	ibi	Miraculous-Life
Period	Apr. 2014 – Aug. 2014	Jul. 2014 – Aug. 2014	Jan. 2016 – Jul. 2016
Participants	46	8	/
Sex F/M	33/13	4/4	/
Average age	82.8	68.9	/
User groups	Older adults, younger & active seniors	Older adults	/
Methods	living lab setting	Log files, media diary, interview	/
Devices	Smartphone, PC, tablet, TV	Smartphone, tablet, PC, TV	/

three projects in a natural setting. Table 3 presents the evaluation settings of the AALuis and ibi project during the field trials.

### 3 Evaluation Results

Although all three prototypes use avatar based interaction technology, the main goal of each project is different. However, the presented projects also share various numbers of comparable aspects. The following section focuses on these comparable aspects but also on project specific findings.

#### 3.1 Evaluation Results from the Concept Phase

##### Project Specific Findings

*AALuis.* The group of healthy and active seniors assessed early mock-ups of the AALuis communication service using a laptop and a smartphone. In general, participants reported that the usage of the service was easy for them. One person noticed that it would be interesting if appointments would synchronize automatically with the calendar on their smartphone or PC. There was also the fear to get too dependent on a mobile device in terms of being available all the time.

Older adults evaluated mock-ups of the AALuis reminder service using a tablet and a TV in an assisted living center where we set up a mobile usability lab. For both versions, we could hardly detect any clear usability issue. Nevertheless, some participants had problems interacting with the prototypes. One reason was that the participants were not familiar with the used remote control. The buttons were too small and it was perceived as too complex. Furthermore, participants were hesitant to touch the tablet and had problems to activate a button. They were too hastily or pressed too hard. Sometimes the system did not recognize these touches and the supervisor guided their fingers by hand. It was not possible to test the avatar in this early stage because the development of this module was still ongoing.

*ibi*. In this early phase, the *ibi* system was represented by two functional mock-ups implemented on the tablet and on the TV. Users liked the simple designed GUI containing three areas (avatar area on the top, textual representation of the spoken dialog in the middle and the interaction area with maximal three buttons on the bottom). Participants emphasized that it was very helpful to see the textual representation of the spoken dialog. These users preferred to read the presented content instead of listening to the avatar. The interaction using the four colored buttons on the TV remote control was easy and intuitive enough for almost all participants.

The used speech recognition model, which was originally designed for seniors using the Styrian dialect, was also appropriate enough for the testing group of Viennese citizens. Participants perceived the vocal interaction with the system as something natural. Users emphasized that they liked the speech recognition because they would not need to put their glasses on to be able to read a message from their relatives.

Investigation on the avatar appearance showed that participants prefer a human like middle-aged female avatar which is dressed in everyday clothes and which interacts in a private surrounding like a living room.

*Miraculous-Life*. Older adults as well as formal and informal caregivers assessed the basic functionalities and avatar mock-ups early on in the project. Based on the evaluation results, a first prototype with reduced functionality was tested by experts on a tablet. Avatar mock-ups in conjunction with a questionnaire revealed that the avatar should act as a friendly personal assistant. Furthermore, most users preferred a human like avatar with a young woman's appearance. The avatar should indicate different interaction modes like waiting, listening and talking.

The expert group analyzed the avatar interaction and the available services. It was found that the dialogues used with the avatar should be more friendly and natural when addressing the older adults. The avatar was considered to have a clearly understandable speech output. Some issues were reported on behalf of wrong pronunciation and intonation of certain words and letters, especially when using the French language. Furthermore, strange system behavior was observed where the avatar was arbitrarily giving commands to himself. This happened because the speech recognition module recognized certain commands given by the avatar while presenting new information.

Apart from the avatar analysis, the experts noticed some inconsistencies in the service workflows, especially regarding the navigation. Additionally, experts found that sometimes the system response time after issuing a command was too long.

## **Summary of the Evaluation Results**

Table 4 provides a comparison of the evaluation results during the concept phase using rating scales. Ratings were performed by project experts based on the empirical findings from the user involvements.

## **Implications for the First Prototype**

*AALuis*. At the smartphone version targeting the healthy older adults, some interaction elements like radio buttons and tabs were too close together and the navigation area should be visible all the time. Some older adults struggled with using the tablet because they used a touch screen for the first time and did not dare to really touch the screen.

**Table 4.** Overview of evaluation results from the concept phase. Rating scale interpretation: Entries on the left side of the vertical line indicate usability or understanding problems, entries on the right side indicate better usability and understanding.

	AALuis	ibi	Miraculous-Life
GUI usability	---   -x-	---   -x-	--x   ---
Service understanding	---   x--	--x   ---	---   x--
Interaction on the smartphone	---   x--	not tested	not applicable
Interaction on the tablet	---   -x-	---x---	---   -x-
Interaction on the TV	---   x--	---   -x-	not applicable
Speech recognition	not applicable	---   -x-	---   x--
Avatar appearance	not tested	x--   ---	---   --x

Therefore, a training period for this target group had to be considered. The standard remote control for the TV version confused some participants due to the high amount of unnecessary buttons, thus later on a simplified version was used.

*Ibi.* Intended services for the first prototype needed to be reconsidered based on the questionnaire results performed in this phase. Not all users were familiar with the touch based interaction so they required a training period. The presented avatar also had to be redesigned, because users preferred rather a female, middle-aged avatar.

*Miraculous-Life.* The avatars text pronunciation and intonation had to be improved so that the older adults do not have problems understanding the avatar. Moreover, the speech recognition had to be disabled during the avatar playback in order to prevent the avatar from activating actions through spoken commands. The service workflow needed also improvements to yield a consistent flow of information and navigation.

### 3.2 Evaluation Results from the Lab Trials

#### Project Specific Findings

*AALuis.* The group of healthy and active seniors evaluated the first functional prototype of the AALuis communication service using a PC and a smartphone. In the beginning, 4 of 5 participants felt more comfortable using the application on the PC, but 3 of 5 participants got used to operate the smartphone rather quickly. Nevertheless, when the participants were asked which device they would prefer to use, 4 of 5 favored the PC. Despite some usability problems, participants gave positive feedback to the well-structured and similar layout of the application on both devices.

The participants in need of care assessed the AALuis reminder service on TV and a tablet. In general, usability of the AALuis reminder service was restricted since all

participants had general problems with the TV navigation and were not familiar using a touch-based device. Since participants had no experience with touch screens and mostly refused to use them, 3 of 4 participants preferred using the TV compared to the tablet.

*ibi.* Although the speech recognition was not implemented into the first prototype, some of the users intuitively tried to respond vocally on questions asked by the avatar. After a short clarification and some hints about interaction possibilities, most of the users were able to confirm the presented dialogs by using the TV remote control. Two users felt a little bit uncomfortable and insecure, hence they refused the interaction with the remote control.

Care receivers had problems to confirm messages on the tablet. They were not used to tablets and were not familiar with the concept of touchable buttons. The device was also not sensible enough and users had to click buttons multiple times in order to trigger an action. Informal and formal caregivers had troubles with the smartphone, trying to activate the on-screen keyboard. Additionally, typing on the small screen was perceived to be a difficult task.

*Miraculous-Life.* The overall impression of the system was positive and the end users found the provided services useful. Two users suggested that the *Miraculous-Life* system could be an interesting solution for older adults who live alone at home. On the contrary, other end users suggested that the VSP is a “tricky solution”, since it could potentially encourage social isolation.

While the avatars speed of speech was adjustable, a similar feature for adjusting the volume was missed by the end users. Furthermore, the end users missed the possibility to skip an avatar video. Many participants did not know when the system accepted the next command and wondered why the system would not react on their behalf. Additionally, the users found that the speech recognition was not always accurate, which lead to an unexpected system behavior.

Two older adults experienced difficulties hearing and understanding the avatar. Furthermore, the gestures of the avatar were perceived to be fluent and fit to the spoken words but the repertoire was too limited to simulate a natural interaction.

Regarding the user interface, older adults stated that the buttons and labels were too small to read. Additionally, the avatar was considered to be too small in the service view and too big in the main view.

### **Summary of the Evaluation Results**

Table 5 presents an overview of the evaluation results from the lab trials using rating scales.

### **Implications for the Second Prototype**

*AALuis.* Evaluation results from the second phase indicated that the mock-ups were easier to use by the participants than the prototypes. Thus, for the final prototypes the automatically generated user interfaces needed to be closer to the mock-ups taking the findings of both reported evaluations into account. While some aspects of the prototypes should be solved with better user interface templates (e.g., better highlighting of the focused UI element), other aspects required updates within the underlying

**Table 5.** Overview of evaluation results from the lab phase

	AALuis	ibi	Miraculous-Life
GUI usability	-x-   ---	---   -x-	-x-   ---
Service understanding	--x   ---	--x   ---	---   x--
Interaction on the smartphone	-x-   ---	-x-   ---	not applicable
Interaction on the tablet	--x   ---	x--   ---	---   x--
Interaction on the TV	x--   ---	-x-   ---	not applicable
Interaction on the laptop	-x-   ---	not tested	not applicable
Speech recognition	not applicable	not tested	---x---

middleware layer (e.g., the differentiation between headline and normal text). The creation of a UI template for each device was considered to be a solution to reduce some of the device-specific usability issues.

*ibi.* The usage of a remote control beyond the TV context was for some users unnatural. They needed time to get familiar with this interaction type. Additionally, the remote control had to be simple and clearly designed in order to increase user acceptance. High quality and more sensible tablets were required. The application for the informal caregivers had to be redesigned. Predefined text blocks should help to eliminate the need for typing and to speed up the message generation process.

*Miraculous-Life.* Since the Miraculous-Life project is currently under development, the following improvements have to be implemented in the final prototype. The UI has to be redesigned by means of avatar size, element contrast and coloring to meet the user requirements and suggestions. The usage of external speakers to better understand the avatar rather than using the built-in ones needs to be assessed further. The current interaction mode of the avatar must be clearly visible and perceivable. Furthermore, a possibility to skip an avatar video is desired. The speech volume must be adjustable directly from the UI by spoken commands. Finally, the system latency has to be reduced significantly.

### 3.3 Evaluation Results from the Field Trials

#### Project Specific Findings

*AALuis.* We found that many participants of both user groups appreciated the idea and the services offered by AALuis. However, using the prototype was inferred from time to time by technical problems. Accordingly, the user experience results did not reflect the positive attitude towards AALuis. While the healthy and active older adults got first of all frustrated, the persons in need of care became more insecure and sometimes did not dare to use the service anymore. Nevertheless, many participants uttered their

regrets that they could not use the services more often to benefit from their offers. Positive experiences of users which faced only a few technical problems suggest that a fully functional product offering these services might be interesting for the two involved user groups.

*ibi*. In general, all participants were satisfied with the functionality and the benefits of the offered services. In some cases the internet connection was not as stable as expected, therefore some messages were delivered with an evident delay.

Users appreciated the possibility to interact with the system using multiple devices. Hence, many users combined these devices, e.g., by starting the dialog on the tablet and continuing the interaction by using the avatar-based TV output and the speech recognition. The overall impression of the speech recognition was quite positive. Unfortunately, a few numbers of dialogs were automatically confirmed by the continuous activated speech recognition. Sources of noise, like the radio speech, triggered commands on the system without explicit intention from the users.

Informal and formal caregivers reported that it was sometimes difficult to use the smartphone because of the small screen size. Furthermore, the problem with the insensitive tablet devices remained. It was not possible to purchase new devices for the field trials so the lab trial devices had to be reused. Informal caregiver remarked that it would be practical to have an automatic synchronization between the *ibi* calendar and a private calendar, e.g., on the smartphone.

### Summary of the Evaluation Results

Table 6 summarizes the evaluation results during the field phase using rating scales.

### Implications for Further Improvements

*AALuis*. Developers have to make sure that the technical setup and the system performance can be guaranteed in terms of stability of the integrated system, functionality of all components, and suitability of the external conditions such as sufficient network

**Table 6.** Overview of evaluation results from the field trials

	AALuis	ibi	Miraculous-Life <sup>a</sup>
Overall usability	---   -x-	---   -x-	/
Service satisfaction	---   -x-	---   -x-	/
System reliability	---   x--	---   -x-	/
Interaction on the smartphone	--x   ---	--x   ---	/
Interaction on the tablet	---   -x-	-x-   ---	/
Interaction on the TV	---   x--	---   -x-	/
Interaction on the laptop	---   x--	---x---	/
Speech recognition	not applicable	---   -x-	/

<sup>a</sup> The Miraculous-Life project has not been evaluated in the field so far.

coverage. Apart from these technical issues, the avatar was well appreciated, especially by the older adults. It would be good to offer more control possibilities to the users in terms of replaying, pausing or stopping the avatar. Other AALuis UI elements would benefit from a more visually appealing graphic design which also requires adjustments in the middleware layer.

*ibi*. The speech recognition must not be activated all the time but just by special activation commands, e.g., by the name of the used avatar. Formal caregivers, who mainly used the smartphone, need a smarter solution to announce a delay of the home-care visit. Neither the typing of a delay message, nor the selection of a predefined text-message are appropriate options for this task on small smartphone screens. As an improvement one can, e.g., use location based services and a scheduler to automatically pre select the user and the message for a delay announcement. Older adults require devices with an acceptable touch sensibility. Therefore, touchable devices have to be pretested before the deployment. For the informal caregivers, automatic calendar synchronization could be a practical feature.

## 4 Conclusion

In this paper, we present evaluation results of avatar-based supporting systems developed in three different Ambient Assisted Living projects. Prototypes developed in the projects AALuis and *ibi* generally use an avatar but the focus is rather on a versatile interaction with different modalities on different devices. In contrast, *Miraculous-Life* strongly focuses on the avatar as a virtual interaction partner and not on modality and device diversity. However, results from all three projects showed that avatar-based interaction in the Ambient Assisted Living context is very well applicable. This kind of interaction, especially when combined with speech recognition, offers big advantages for the target group.

**Acknowledgments.** The project AALuis was co-funded by the AAL Joint Programme (REF. AAL-2010-3-070) and the following National Authorities and R&D programs in Austria, Germany and The Netherlands: BMVIT, program benefit, FFG (AT), BMBF (DE) and ZonMw (NL).

The *ibi* project was co-funded by the benefit programme of the Federal Ministry for Transport, Innovation and Technology (BMVIT) of Austria.

The *Miraculous-Life* project is co-funded by the European Commission under the 7th Framework Programme (Grant Agreement No 611421).

## References

1. Mayer, C., Morandell, M., Gira, M., Sili, M., Petzold, M., Fagel, S., Schüler, C., Bobeth, J., Schmehl, S.: User interfaces for older adults. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2013, Part II. LNCS, vol. 8010, pp. 142–150. Springer, Heidelberg (2013)
2. Sili, M., Mayer, C., Morandell, M., Gira, M., Petzold, M.: A practical solution for the automatic generation of user interfaces – what are the benefits of a practical solution for the automatic generation of user interfaces? In: Kurosu, M. (ed.) HCI 2014, Part I. LNCS, vol. 8510, pp. 445–456. Springer, Heidelberg (2014)

3. Mayer, C., Zimmermann, G., Grguric, A., Alexandersson, J., Sili, M., Strobbe, C.: A comparative study of systems for the design of flexible user interfaces. *J. Ambient Intell. Smart Environ.* (2014, in press)
4. Mayer, C., Morandell, M., Gira, M., Hackbarth, K., Petzold, M., Fagel, S.: AALuis, a user interface layer that brings device independence to users of AAL systems. In: Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W. (eds.) *ICCHP 2012, Part I. LNCS*, vol. 7382, pp. 650–657. Springer, Heidelberg (2012)
5. Hanke, S., Sandner, E., Sili, M., Hochgatterer, A., Ben Moussa, M., Christodoulou, E., Trindadem, P., Samara, G., Andreou, P., Wingskölgén, C., Van Der Aa, N., Stockloew, C.: Virtual support partner: a real-time, emphatic elder care system that attends to the daily activity and safety needs of the elder at home, during his normal daily life. In: *Broader, Bigger, Better – AAL solutions for Europe - Proceedings of the AAL Forum 2014 Bucharest* (in press)
6. Preece, J., Rogers, Y., Sharp, H.: *Interaction Design: Beyond Human-Computer Interaction*. Wiley, New York (2011)
7. Nielsen, J., Clemmensen, T., Yssing, C.: Getting access to what goes on in people’s heads? - reflections on the think-aloud technique. In: *NordiCHI 2002* (2002)
8. Lewis, C., Rieman, J.: *Task-Centered User Interface Design: A Practical Introduction*. University of Colorado, Boulder (1993)
9. Brooke, J.: SUS: a quick and dirty usability scale. In: Jordan, P.W., Weerdmeester, B., Thomas, A., McLelland, I.L. (eds.) *Usability Evaluation in Industry*. Taylor & Francis, London (1996)
10. Wharton, C., Rieman, J., Lewis, C., Polson, P.: The cognitive walkthrough method: a practitioner’s guide. In: Nielsen, J., Mack, R.L. (eds.) *Usability Inspection Methods*, pp. 105–140. Wiley, New York (1994)
11. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering people*. ACM (1990)