Wir im Kiez

Multimodal App for Mutual Help Among Elderly Neighbours

Sven Schmeier
German Research Center for Artificial Intelligence (DFKI)
Alt-Moabit 91c
10559 Berlin, Germany
sven.schmeier@dfki.de

Aaron Russ
German Research Center for Artificial Intelligence (DFKI)
Alt-Moabit 91c
10559 Berlin, Germany
aaron.russ@dfki.de

Norbert Reithinger
German Research Center for Artificial Intelligence (DFKI)
Alt-Moabit 91c
10559 Berlin, Germany
norbert.reithinger@dfki.de

ABSTRACT

Elderly people often need support in everyday situations – e.g. common daily life activities like taking care of house and garden, or caring for an animal are often not possible without a larger support circle. However, especially in larger western cities, local social networks may not be very tight, friends may have moved away or died, and the traditional support structures found in so-called multi-generational families do not exist anymore. As a result, the quality of life for elderly people suffers crucially. On the other hand, people from the broader neighborhood would often gladly help and respond quickly. With the project *Wir im Kiez* we developed and tested a multimodal social network app equipped with a conversational interface that addresses these issues. In the demonstration, we especially focus on the needs and restrictions of seniors, both in their physical and psychological limitations.

Keywords

mobile multimodal, usability, conversational interface, elderly

1. INTRODUCTION

In the project *Wir im Kiez*, funded by the German ministry of education and research (BMBF), we developed together with the industrial company Cocomore AG[1] a social media platform which brings people in the neighborhood together to help each other in daily activities. It is intended to facilitate a social network which brings people in real life in contact within the local area, keep in touch with friends and family members, or just get informed about the latest local events. The project especially targets elderly people to become part of this community, notably those who lost their marriage partner or friends, who cannot draw on support from family and friends, or who are just feeling lonely for whatever reason. As central issue in the interface design process we had to consider age-related characteristics of the intended user group. With increasing age, typical ailments comprise loss of visual acuity, reduced perception of color-contrasts (e.g. worse for blue and green, better red or yellow), decreased ability to adjust changes in illumination, reduced field of vision, hearing loss especially for high pitched sounds, fine-motor control decline, and also loss of cognitive functions, as for example a decreased short-term memory. These issues play an important role when considering age-related effects in UI design. During the project, feedback from the target group has been collected frequently in usability tests with mock-ups and intermediate system versions. This feedback has been included in the design and realization to ensure that all important aspects and limitations of the target users group were addressed. The UI design is also guided by experiences with previous projects in this area [1]. The application itself is designed for tablet devices. Tests have shown that the target group can handle input via touch especially well compared to input via mouse and keyboard. Feedback mechanisms also play an important role while navigating the application. In Fig.1 we see the home screen of the application; note the large

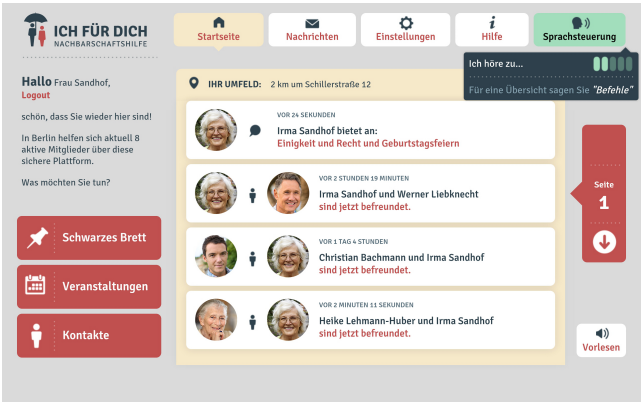

Figure 1: The home screen with active speech control (dt. "Sprachsteuerung"); original size 10.1 in /25.7 cm. *Demonstrator available as Chrome Extension at* https://chrome.google.com/webstore/detail/aeibgffcbkeeipgoeogbehkdickahmdf

[1] Cocomore AG, Gutleutstr. 30, 60329 Frankfurt am Main, Germany

ICMI 2015, November 9–13, 2015, Seattle, WA, USA.
ACM 978-1-4503-3912-4/15/11.
DOI: http://dx.doi.org/10.1145/2818346.2823300 .

touch targets and the contrast of the chosen colors. In addition to guarantee a barrier-free user experience we added a *conversational interface*. The user can navigate through the interface by using natural language dialogs. A user may talk to the system in the way he/she wants to: no commands need to be learned or recognized.

2. TECHNOLOGY

The multimodal platform is realized using the Mobile Multimodal Interaction and Rendering (MMIR) framework[2], a framework for building lightweight dialog systems on basis of web technologies. Combined with Cordova[3], the framework allows targeting the major mobile platforms. MMIR provides several modules to add speech functionality like Automatic Speech Recognition (ASR) and Text To Speech (TTS) synthesis for different platforms. The Android version uses the Nuance NDEV solution for ASR and TTS, whereas the PC app –in form of an extension for the Chrome browser– utilizes the built-in Web Speech API of Chrome for ASR, and an OpenMARY[4] server for TTS over the web. The content pages are realized as standard web pages that are enriched with our multimodal dialog technology.

When using the conversational interface, i.e. speech control, the recognition results of the ASR are classified according to anchor points of the site map. The classification algorithm that we employ is based on a Rocchio based Machine Learning approach and has been proven to work extremely well on small and unbalanced training data [2]. The algorithm is trained continuously by the users' utterances collected during the use of the system. In case of ambiguities during the categorization, the dialog system will ask questions for clarification (cf. Fig. 2). The dialog system's processing of the event-driven input and application states is realized using State Chart XML (SCXML). It tracks not only the spoken user events, but also the direct tactile user input on the buttons of the page to keep track of the interaction state, to create haptic feedback. In addition to the command-and-control oriented navigation, the dialog system also enables the user to dictate –and correct– text input, e.g. for creating messages and help-offers or -requests.

3. DEMONSTRATION DESCRIPTION

As mentioned before users can interact with system with the modalities GUI/touch and speech. System feedback comprises sounds/earcons and haptics/vibration for confirming interaction. Especially haptic feedback is very important for the targeted user group. Once the conversational modus, i.e. speech control, is activated, the system listens continually for spoken commands, until it is explicitly deactivated again. This modus is prominently indicated by showing a popup-dialog (see upper right in Fig. 1). It contains a 5-bar visual feedback for the microphone input as well as a hint for listing available speech command: the listing can be triggered and used either by voicing the hint, or by touch.

In context of the CARE classification [3], interactions with the dialog system are mainly multimodally *Equivalent*, i.e. an action can be equivalently triggered either by a voice or haptic interaction(s); this allows users to choose their preferred way for controlling the system. Some actions that can

[2] https://github.com/mmig/mmir
[3] https://cordova.apache.org
[4] https://github.com/marytts

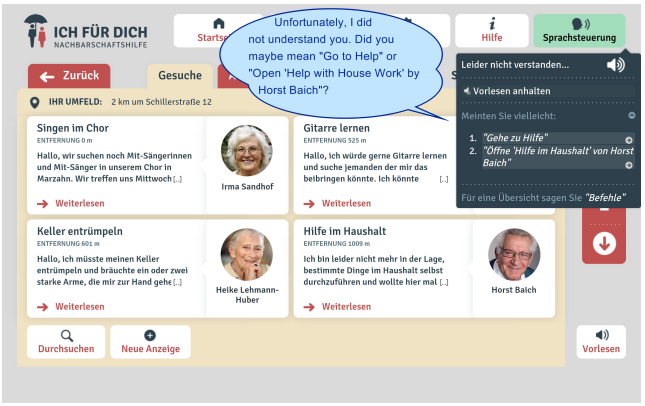

Figure 2: Example for disambiguation: candidates for likely commands are read out via speech synthesis, and also shown in the speech control dialog as *touchable* entries.

be triggered using speech commands may not be available by touch in the current view. In this case, multimodal *Equivalence* is provided by way of the command listing, which can be opened in the command popup-dialog. The currently available commands are listed and can be triggered by touch. Some few actions can only be used in one singular modality (*Assignment*), namely starting the speech control and stopping TTS[5] can only be accomplished using touch. Also user authentication –entering the user name and password– is not possible via speech.

With respect to the targeted user group, the interactions are designed to be simple. This is the main reason, that the dialog system does not provide *Redundant* multimodal interactions: transferring information redundantly in multiple modalities would result in one interaction replacing the other, i.e. *Equivalence* of the interactions. For the same reasons, the system does not provide *Complementarity* multimodal interactions.

4. ACKNOWLEDGMENTS

The work presented in here was funded by the German Ministry of Education and Research in the project *WirImKiez* (16SV6323). The authors would like to thank the Cocomore team and especially Andreas Aue and Nico Beck for their contributions to this work.

5. REFERENCES

[1] M. Haesner, K. Schumacher, A. Steinert, E. Steinhagen-Thiessen, and M. Goevercin, "Evaluating an online cognitive training platform for older adults - user experience and implementation requirements," *Journal of Gerontological Nursing*, 2015.

[2] S. Schmeier, *Exploratory Search on Mobile Devices*. Saarbruecken, Germany: DFKI-LT - Dissertation Series, 2013.

[3] L. Nigay and J. Coutaz, "Multifeature systems: The care properties and their impact on software design," *Intelligence and multimodality in multimedia interfaces*, 1997.

[5] Barge-in for speech input is not supported