Speech in Multimodal Interaction: Exploring Modality Preferences of Hearing Impaired Older Adults

Stephanie Schwarz
AIT Austrian Institute of Technology GmbH
1210 Vienna, Austria
stephanie.schwarz@ait.ac.at

Diotima Bertel
AIT Austrian Institute of Technology GmbH
1210 Vienna, Austria
diotima.bertel@univie.ac.at

Julia Himmelsbach
AIT Austrian Institute of Technology GmbH
1210 Vienna, Austria
julia.himmelsbach@ait.ac.at

Manfred Tscheligi
AIT Austrian Institute of Technology GmbH
1210 Vienna, Austria
manfred.tscheligi@ait.ac.at

Ulrich Lehner
AIT Austrian Institute of Technology GmbH
1210 Vienna, Austria
ulrich.lehner@ait.ac.at

Abstract

This paper addresses usability challenges of speech and multimodal interfaces by investigating modality preferences and interaction quality of a mobile medication in-take reminder. Older users (normal hearing, hearing impairment) interacted with input and output modality conditions on a smartphone in a laboratory study. Interestingly hearing impaired users prioritized voice only interaction over other modalities, which clearly points up the relevance of speech interaction for people with hearing impairment.

Author Keywords

Multimodal interaction; speech-based interaction; hearing impairment; older adults; interaction quality.

ACM Classification Keywords

• Human-centered computing~Natural language interfaces

Introduction

Speech interaction has been successfully added as both input and output in m-health applications e.g. for older users using a mobile medication reminder [6]. While unimodal speech interaction has shown to be accepted by older adults in selected contexts [8], multimodality

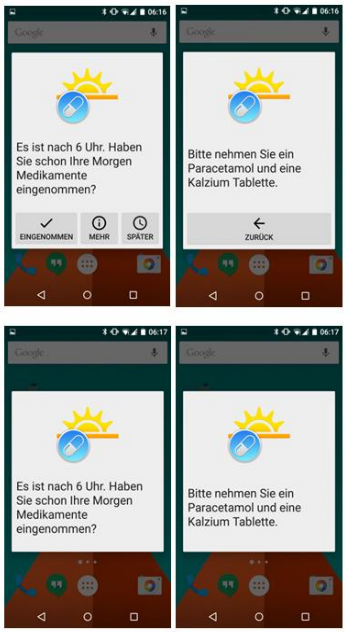

Figure 1**:** Medication reminder GUI for touch input in the conditions GUI, VOG, VG (top) and for speech input in the conditions VGO (bottom). Speech-only condition (VUI) was provided without any graphical output

including speech might compensate for some of its dedicated shortcomings.

Beside evidence that modality choices are influenced by the given task context [2], previous research revealed that modality preferences might be influenced by individual characteristics. Effects of gender, but not of age (younger and older adults) on modality preferences were found when offering unimodal and multimodal input possibilities including voice, free-hand gesture, touch screen [7]. However there is a lack of knowledge about whether and how speech interaction is used by users with specific physical constraints such as age-related hearing loss. Recent work investigated modality choices of older adults when using mobile applications via speech and touch interaction over multiple weeks [4]. According to their results hearing impairment still represents a major barrier for the use of speech.

We suppose that the combination of a hearing aid device for audio output and a mounted microphone with dedicated smartphone applications providing voice interfaces (VUI) might be beneficial for hearing impaired users to interact with mobile services in different contexts. Based on this approach and the results of a previously conducted requirements analysis [1], a smartphone-based medication reminder application enabling voice input and output was developed as part of the AHEAD project[1]. In this work we investigate the relationship between a specific user characteristic (hearing impairment) and multimodal system characteristics enabling speech and touch interaction (interaction quality). Our aim is to identify modality preferences of older adults with and without hearing impairment, as well as to learn how the latter assess the interaction quality of different VUI and GUI modality combinations. Older adults were invited to evaluate the usability and perceived interaction quality of a mobile medication in-take reminder application offering unimodal and multimodal interactions (input, output) via speech and touch.

Apparatus

Standard Bluetooth headphones with a command button connected to a smartphone application as shown in Figure 2 were used to provide visual and auditory reminders. The prototype was built on the Android platform running on a Nexus 5 device, capable of operating in the background. Standard Android API functionality, namely text-to-speech (TTS) service and speech recognition for the spoken user input was used. Based on requirements analyses a synthesized voice was used for TTS output [1]. The prototype simulated the in-take reminder by following a Wizard-of-Oz approach. Reminders were initiated by sending a text message to the test device containing a code to trigger the modality. This approach was robust without the need of any internet connection.

Input and Output Modality Combinations

The reminder on the phone asked whether the scheduled pills had been taken. The participant could respond to the dialogue via voice command or button press e.g. saying or taping the words i) "yes" or "taken" to confirm the in-take, which was acknowledged with a short reminder on the next medication in-take, ii) "more" or "which" to hear further information, or iii) "no" or "later" to postpone the in-take. For voice input, in case automatic speech recognition (ASR) failed, a voice message asked to repeat the command ("I did

[1] www.ahead-project.eu

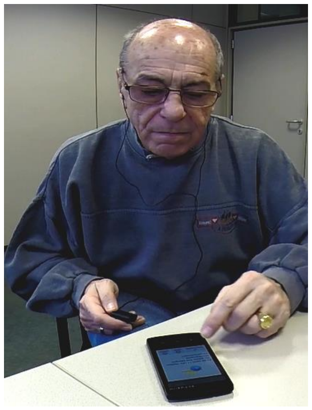

Figure 2: Participant replying to the TTR reminder via touch (VOG)

not understand you correctly", repetition of the question), including hints on how to correctly respond to the system. The dialogue could be accessed by the modalities of touch and/or speech in various unimodal and multimodal conditions (see Table 1). 5 modality conditions were presented to participants in counter-balanced order to avoid position effects.

For unimodal interaction in the VUI condition the dialogue was completely speech-based (no GUI). All messages were synthesized and the user replied with voice commands. In the GUI condition the interaction was limited to a graphical user interface with textual descriptions and buttons on the smartphone (see Figure 1). Concerning multimodal conditions, during VGO all interactions were voice-based (TTR messages and speech input); additionally, the system output was displayed on the device. Concerning multimodal conditions, during VGO all interactions were voice-based (TTR messages and speech input); additionally, the system output was displayed on the device. VOG interaction was touch-based with visual output, and displayed text was read by the synthetic voice when it appeared.

	Unimodal	**Multimodal**
GUI input and output	Touch interaction (GUI)	Touch interaction, speech output only (VOG)
VUI input and output	Speech interaction (VUI)	Speech interaction, graphical output only (VGO)
GUI, VUI input and output	-	Completely synced touch and speech interaction (VG)

Table 1: Modality combinations and conditions

The VG condition was presented combining all interaction possibilities. Participants were presented with spoken and textual messages and could respond either with voice commands or by pressing an on-screen button. By this approach, we achieved a comparison of all possible combinations of touch and speech interaction.

Procedure

After a short introduction the facilitator explained the handling of the head set for performing speech commands and the reminder prototype. Volume level for the TTS messages was adjusted according to participants' needs. The main part of the study consisted of listening and replying to the in-take reminders via the modality combinations as shown in Table 1. Participants gave verbal feedback to gather in situ insights regarding the interaction quality and usability. A semi-structured interview was conducted focusing on participants´ general opinion. Qualitative data was analyzed with an inductive content analysis.

A questionnaire was filled in to subjectively assess the perceived usability of the reminder [3]. Finally, participants assessed perceived dialogue and interaction quality (adapted from [5]) and ranked all modality conditions according to personal preferences.

In total 18 older adults (14 male, 4 female) aged between 57 and 80 years (mean=66.2, sd= 8.1) took part. 9 participants (50%) had a hearing impairment. 7 out of these 9 persons had a diagnosed hearing impairment and 4 used hearing aids. 11 participants stated that they were accustomed to taking daily medicine (varying intake routines).

Results

Usability feedback revealed homogeneously positive ratings by both hearing impaired and users with normal hearing regarding the tool's attractiveness, efficiency, stimulation and novelty (see Figure 3). All users perceived the conversational aspect of the reminder as holistic and fluent, while the course of the dialogue was perceived as clear and logical. The information was complete except for the fact that the system did not disclose the exact time of the next reminder in case a pill had not been taken yet. Dialogue utterances were perceived as well timed, although some participants would have wished to receive additional information about the medication in question (med details, picture of the package). Overall, assessments of interaction quality did not differ between older adults with normal hearing capacities compared to hearing impaired ones.

Modality Preferences

Conditions were experienced differently (see Figure 4). Surprisingly for unimodal interaction, hearing impaired users clearly prioritized VUI interaction while older users without hearing loss preferred classic GUI interaction over VUI. Effects of hearing impairment on the ratings of VUI (p= .027) and GUI (p= .002) are significant. Spearman correlation shows that the relationship between hearing impairment and positive ratings for VUI (0.68, p< .01) and GUI (-0.5, p< .05) can be considered as strong. In general, participants anticipated using voice input only in private settings (e.g. at home) seeing potential social and privacy issues occurring when interacting with the system in public. As expected, regarding the multimodal approaches, the synced variant enabling VUI and GUI interaction (VG) was prioritized by all participants over the two other variants (see Figure 4).

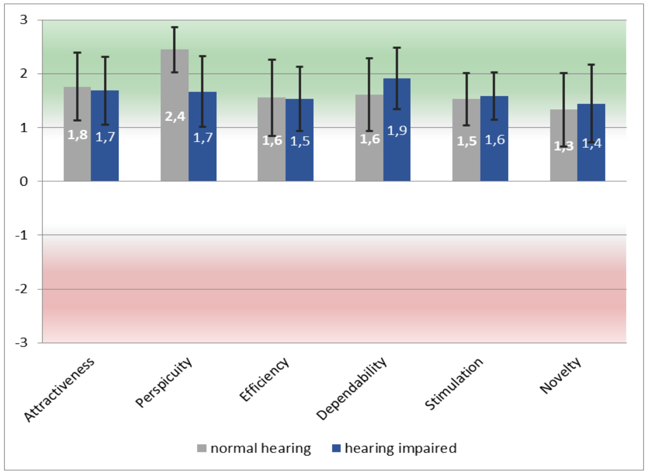

Figure 3: Results of the UEQ questionnaires about the medication reminder application (best: +3, worst: -3)

The VG condition was perceived as being suitable for all situations, also seen as a plausible fallback in case VUI interaction would fail. Although some participants claimed that the approach would require some time to get familiar with, this version exploited the full range of interaction possibilities with the reminder, and in general, this approach was rated to have potential for the future.

Hearing impaired users appreciated the multimodal condition providing GUI interaction and voice output only (VOG). They anticipated the usefulness of such approach in daily life, if provided with a hearing aid or audio transmission device to wear. In case of an audio message being missed, one could rely on the GUI content. Also the VOG condition would avoid privacy issues in public if provided through ear or head phones. However, the distractive character of speech output would remain and cause feelings of discomfort in social settings.

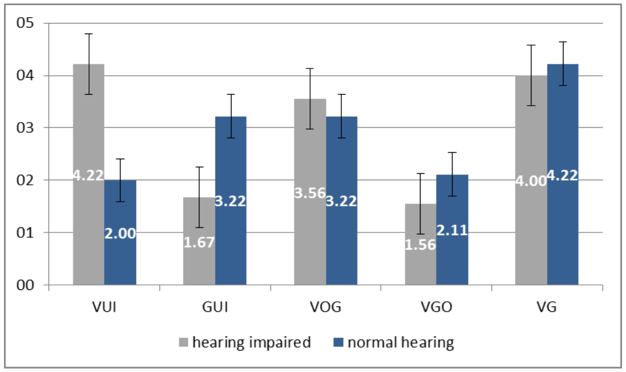

Figure 4: Mean ratings of reported modality (combinations) preferences (1= lowest to 5= highest)

Finally, the multimodal condition for voice interaction and GUI output only (VGO) received the lowest rank by both groups. Although VUI interaction was supported with available visual feedback, no touch entry could be made. Most participants perceived VGO as the most mentally demanding condition; i.e. concentrating on speech input and output with visual output only was too demanding. This approach lacked of intuitiveness for most participants.

In general, unimodal VUI interaction was experienced as easy and effective but not suitable for all situations, e.g. while listening to music, in public and social contexts where voice input would cause discomfort for the user and lead to potential privacy issues. The advantages of voice output combined with GUI interaction (VOG) were appreciated, especially when anticipating the use of some wearable device for audio output. This condition avoids feelings of frustration in case of ASR inaccuracies while at the same time it does not require having the smartphone at hand for information reception.

Discussion and Future Work

During speech interaction (unimodal, multimodal) participants enjoyed the conversation with the system and were engaged by dialogues with the synthesized voice. The overall concept was clear while information content was comprehensive and almost complete. While voice interaction only (VUI) had some advantages such as hands free interaction, real time notification through the additional GUI (VOG, VG) kept its relevance as fallback (also due to the users' general skepticism towards ASR quality). Hence, user feedback supports previous findings that VUI only is not suitable for all contexts [8], but it might complement the interaction when used in a multimodal design.

Overall benefits of multimodality were found as the synced combination of VUI and GUI interaction outperformed unimodal modality conditions regarding subjective preference ratings by participants. For future interaction design, we therefore recommend focusing on complete and synced multimodality in contrast to implementing additional modalities for only input or output. Future work might put effort toward adaptive interfaces, personalizing input and output according to user characteristics, individual preferences and context.

Findings further reveal that hearing impaired users preferred speech only interaction over touch only. A possible explanation for this interesting result is that hearing impaired users might be trained to cope with ambiguous audio stimuli. They might be used to put more attentional efforts into understanding audio stimuli. Similarly the positive connotation of hearing aids as support in everyday life might have effects on the acceptability of speech interaction. Studies reveal the relevance of previous experience with ICT and speech interaction as a factor of modality choices [e.g. 4]. Wearing a hearing aid or even the anticipation could also be seen in this context. Hence, (anticipated) experience with audio information enabled through a device, i.e. the hearing aid, might be beneficial for the acceptance of and preference for speech interaction.

Previous work consistently portrays hearing impairment as a barrier for speech interaction. Our results question this portrayal. More work is needed to clarify the relationship between user characteristics and speech interaction: individual needs, perceptual capacities and previous experience with speech and hearing devices need to be considered in future research and design.

Acknowledgements

This research was conducted as part of the project AHEAD (AAL-2012-2-167), partially funded by the European Ambient Assisted Living Joint Programme and the National Funding agency in Austria.

References

1. Pierre Barralon, Ivo R. Maia Martins, Nicole Merkle, Stephanie Schwarz. Augmented Hearing for elderly people: From requirements to implementation. *In Proc. of International Conference on Information and Communication Technologies for Ageing Well and e-Health*, 2015, 59-66, Lisbon, Portugal.
2. Sanjay Ghosh and Joshi Anirudha. Exploration of Usage Patterns for Multimodal Input Interactions. *In Proc. of the India HCI 2014 Conference on Human Computer Interaction*, ACM, 2014.
3. Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. *In Proc. of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society on HCI and Usability for Education and Work* (USAB '08), Andreas Holzinger (Ed.). Springer-Verlag, Berlin, Heidelberg, 63-76.
4. Julia Himmelsbach, Markus Garschall, Sebastian Egger, Susanne Steffek, and Manfred Tscheligi. 2015. Enabling accessibility through multimodality? interaction modality choices of older adults. In *Proc. of the 14th International Conference on Mobile and Ubiquitous Multimedia* (MUM '15). ACM, New York, NY, USA, 195-199.
5. ITU-T Rec. P.851, Subjective Quality Evaluation of Telephone Services Based on Spoken Dialogue Systems, International Telecommunication Union, Geneva, 2003.
6. Antonio Teixeira, Flora Ferreira, Nuno Almeida, Ana Rosa, Jose Casimiro, S. Silva, A. Queirós, and André Oliveira, "Multimodality and Adaptation for an Enhanced Mobile Medication Assistant for the Elderly," *in Proc. Third Mobile Accessibility Workshop* (MOBACC), CHI 2013, 2013.
7. Benjamin Weiss, Sebastian Moller, and Matthias Schulz. Modality Preferences of Different User Groups. *In Proc. 5th International Conference on Advances in Computer Human Interactions*, (2012), 354-359.
8. Linda Wulf, Markus Garschall, Julia Himmelsbach, and Manfred Tscheligi. 2014. Hands free - care free: elderly people taking advantage of speech-only interaction. In *Proc. of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (NordiCHI '14).