Age-appropriate Participatory Design of a Storytelling Voice Input in the Context of Historytelling

Torben Volkmann, Michael Sengpiel, Rita Karam and Nicole Jochems
Institute of Multimedia and Interactive Systems, University of Lübeck, Ratzeburger Allee 160, Lübeck, Germany
{volkmann, sengpiel, karam, jochems }@imis.uni-luebeck.de

Keywords: Aging Users, Human-centered Design, Participatory Design, Voice Interface.

Abstract: With the demographic change, the percentage of older adults steadily increases. At the same time, new information and communication technologies (ICT) emerge at an ever-increasing rate, making it imperative to consider older adults in the development process to achieve the best possible usability and acceptance for older adults. This paper describes the development of a storytelling input component in the context of Historytelling (HT), which provides a digital interactive platform for older adults to share life stories across generations, potentially improving their health and wellbeing. HT follows the HCD+ (Human Centered Design for Aging) approach, claiming that older adults should be integrated as co-designers throughout the development process. A total of 19 older adults (M=68 years old) participated in 3 studies to analyze, evaluate and design a storytelling voice input, investigating voice communication technology for conversational agents. They were successfully involved in the design process, with methods adjusted to accommodate specific user characteristics of older adults and substantially contributed to the further development of the HT project, exploring the two central research questions regarding the type of voice input suitable for older adults and the minimal requirements for a conversational agent.

1 INTRODUCTION

With the demographic change, the percentage of older adults steadily increases. At the same time, new information and communication technologies (ICT) emerge at an ever-increasing rate, making it imperative to consider older adults in the development process to achieve the best possible accessibility and usability for older adults. Thus, we should value older adults as possible co-designers in the development process (Sengpiel et al., In Press). The Historytelling project (HT) is a research project relying on the strengths of older adults, giving them a tool to tell life stories on a digital platform and share them with other people. HT seeks to have a positive influence on a societal, a group and an individual level. On the societal level, HT fosters multi-perspective historiography, on the group level strengthening of family bonds and friendships and on the individual level a place to actively reminisce and reach out to others. The project addresses these challenges by developing a digital social platform for older adults, giving them the power to record, visualize and share their life stories.

One key aspect of HT is the actual storytelling of older adults. Passing on stories is mostly done via speech as it is the most natural channel and stories are mostly passed on in face to face conversations, having their own research field (Bornat et al., 2015) and potentially positive effects on the listeners (Isbell et al., 2004). Thus, the challenge for HT is to transfer and implement this conversational element to technology in the best possible manner.

Thus, alongside the development of a voice input component for HT, the goal of the research was to explore two research questions: (Q1) Which type of voice input is suitable for older adults? (Q2) What are minimal requirements for a conversational agent for older adults in the context of Historytelling?

An HCD+ (Human Centered Design for Aging) approach focusing on participatory design and consideration of user characteristics was used to answer these research questions and for the actual development of voice input for HT. Hence, older adults took part throughout the development process.

Volkmann, T., Sengpiel, M., Karam, R. and Jochems, N.
Age-appropriate Participatory Design of a Storytelling Voice Input in the Context of Historytelling.
DOI: 10.5220/0007729801040112
In *Proceedings of the 5th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2019)*, pages 104-112
ISBN: 978-989-758-368-1

1.1 State-of-the-art of Voice Communication Technology

As Schafer (1995) pointed out, there are four challenges regarding voice communication technology: "(i) hardware/software implementation of the system. (ii) synthesis for voice output, (iii) speech recognition and understanding voice input, and (iv) usability factors related to how humans interact with machines". Schafer (1995), Cohen and Oviatt (1995) also point out advantages of voice input: Speech is the natural way to communicate; voice is usable even if the hands or eyes are busy; voice communication is accessible for handicapped persons; sometimes natural language interaction is preferred and "pronunciation is the subject matter of computer use" (Cohen and Oviat, 1995).

In the last few years, different digital voice assistants such as Google Home and Amazon's Echo were developed and marketed that increased the overall usage of voice input systems. Thus, the longstanding problem of speech recognition and understanding voice input seems to have been solved for the consumer market, at least in a narrowed context (Hailpern et al., 2010; Hazen et al., 2004; Levin and Lieberman, 2004). Especially new developments of neural networks bring constant improvements to the field of voice recognition (Arik et al., 2017).

Technologically, there are three options to process the voice input: audio recording, speech-to-text input and automatic transcription. Audio recording is possible using various ICT, such as laptops, tablets and smartphones. The speech-to-text input converts spoken words instantaneously into text, whereas the automatic transcription converts the recorded audio to text afterwards and is often used for automatic interview transcription.

1.2 State-of-the-art of Embodied Conversational Agents

With a strong focus on reminiscing and passing on life stories, it is most likely that HT will provoke emotional reactions during the process of telling and listening to stories. Thus, it is important to design an interface that responses to these reactions. One possibility to do so is by using avatars, which can answer to emotional stories via facial expressions and gestures (Sutcliffe, 2017). Using the OCC (Ortny, Clore, Collins) model, Sutcliffe (2017) proposes a taxonomy based on 22 emotions, split into reactions to events, agents (other people) and objects to design suitable reactions of systems.

Integrating these emotions via faces can be done with embodied conversational agents (ECA), which gained attention in the last few years in research (Tsiourti et al., 2014). ECAs are virtual characters, which have the same properties as humans in a face-to-face communication and have been successfully integrated into projects with older adults (Cassell, 2000). It became apparent, that older adults followed instructions by ECAs better than those by classic user interfaces and that they had a subjectively had a positive influence on recall tasks (Ortiz et al., 2007; Tsiourti et al., 2018).

Isbister and Doyle (2002) developed a taxonomy relevant for the development of an ECA. It consists of five different categories to classify and evaluate ECAs: Believability, Social interface, Application domains, agency and computational issues and production.

1.3 Participatory Design Process

Participatory design is often seen as a third space of human computer interaction in which the knowledge of different stakeholders such as the user and the developer can be combined, giving new insights to perform new actions (Muller, 2003). Thus, fundamental design decisions are based on information gathered by involving potential users into the discussion about functionality, features and look-and feel. Participatory design especially helps if the developers are not specialists in the observed field.

There are special demands for participatory design methods involving older adults. For them, some conventional design methods may even be inappropriate (Eisma et al., 2004).

In a literature review, Orso et al., (2015) found that especially visual prompts (graphical representation of an abstract concept), experiencing (giving a direct first-person perspective, i.e. with video sketches), hands on (evoking the reaction and opinion on a tool by providing a physical object instead of a conceptual prototype) and natural tasks (performing a task that is similar to the final context of use) are used when older adults are involved in designing interactive technology. For the HT-development, the HCD+ approach was used, emphasizing the importance of involving the user in every crucial design step as participatory designers. HCD+ especially provides guidelines regarding the recruitment of participants, the atmosphere when working with older adults and required adaptations concerning the concrete execution of methods (Sengpiel et al., In Press).

In the analysis phase, current technological approaches were tested and evaluated by older adults. In the design and conception phase, an experimental game was conducted to develop specific design elements. As a last step, a task-based evaluation of the developed interface was conducted.

2 VOICE INPUT ANALYSIS

2.1 Method

To answer the first research question ("Which type of voice input is suitable for older adults?"), an evaluation of state-of-the-art software was conducted. Thus, the three different input technologies were subjectively evaluated.

Interviews are an important method in an HCD development process, especially in the beginning (Wood, 1997). Due to the potential lack of computer literacy in the group of older adults (Fisk, 2009; Sengpiel and Dittberner, 2008), a task-based evaluation of various technologies was conducted in this initial study.

In the evaluation, eight older adults aged between 60 and 73 (M=67.5, SD=3.7) took part. Four of them were males and four females. They were recruited through personal contacts, mailing lists and notice boards. Seven interviews took place at the university, one took place at home due to physical handicap.

The evaluation was divided into three parts: introduction, practical work, and follow-up. In the introduction, participants introduced themselves and where asked about key aspects of their life and technology usage. In the practical work phase, the older adults got a task for three different input approaches. Google docs was used to demonstrate the speech-to-text capabilities, the software "Speak a Message" was used for audio recording and transcription. Qualitative post-interviews were conducted after every task.

As a follow-up, each participant was asked for their favorite input approach and filled in a questionnaire testing their computer literacy (Sengpiel and Dittberner, 2008) and affinity for technology (Franke et al., 2018)

2.2 Results

In particular, the transcription method was not well known among the participants or they had outdated information on technical possibilities and were positively surprised about the initial quality of the automatic transcription.

All participants stated that an assistive system and better feedback by the software would be appreciated. The preferred feedback varied among the participants, so that visual and auditory assistance should complement each other.

The results show a strong heterogeneity within the group of participants regarding affinity towards technology and computer literacy. Thus, some participants were confident in using the presented software, whereas others needed some time to adjust to the task. Faster participants showed a higher affinity towards technology and computer literacy and stated that they tend to find solutions on their own when problems occur.

All (N=8) participants had either a laptop (6) or a computer (3) at home and used either a smartphone (5) or a cell phone (3). They used computers mainly for word processing, mailing and targeted information searching, with a weekly average time of M=18.9 hours (SD=7). On average, they scored M=20.4 (SD=4.17) on the computer literacy scale (CLS, max = 26), which is still low compared to a younger group (M=23.9), but relatively high compared to other older adults (M=14.4, Sengpiel and Dittberner, 2008). Also, they scored M=2.8 on the affinity for technology interaction scale (ATI, SD=0.9, max score = 6).

The participants stated that their technical difficulties were situational and rather hard to describe. When problems occurred, they would mainly turn to friends or family or seek professional help. Three participants stated that they try to find the solution on their own first. However, they also desire assistance provided by the device itself. Alternatively, integrated tutorials as videos would be appreciated, an approach that has been described by Sengpiel and Wandke (2010) among others. The practical part of the study could only be conducted with 7 of the 8 participants.

2.2.1 Speech-to-Text

Five of the seven participants had never used speech-to-text input, and even the two participants who had used this technology before were surprised by the accuracy of the results.

Three participants stated that the conversion from speech to text was too slow, impairing oral fluency. Also, some problems with speech were ambiguous or not seen at all. The software was not "user friendly", since finding functionality was difficult and it was not clear when the recording had started.

2.2.2 Speech-to-Audio

Six of the seven participants had used a dictation device to record audio before. Foremost, participants liked the simplicity of that method, the possibility to replay and edit the audio footage later, and the fact that audio authentically captures the atmosphere.

2.2.3 Transcription

None of the participants had used audio transcription before, but five out of seven participants liked the possibility to have both, audio and text. Assessed particularly positively was the unobtrusiveness of the method, maintaining the oral fluency. Nonetheless, the quality of the initial transcription is crucial for further adoption.

2.2.4 Preferred Input Method

The overall quality and usability aspects of each method played a big role in participants' preferences. Furthermore, intended audience and purpose are key drivers of the preferred method. If the goal was to write a short story quickly, participants would choose the speech-to-text input. The transcription technology was preferred especially for longer, more meaningful stories. Table 1 shows the acceptance among the participants, multiple answers were possible. Since older adults preferred the transcription technology, it will be used for further development.

Table 1: Acceptance frequencies of input methods among participants (N=8).

Technology	Acceptance frequency
Transcription	6
Audio recording	2
Speech-to-text	4

3 AGE-APPROPRIATE DESIGN

3.1 Methods

To answer the second research question ("What are minimal requirements for a conversational agent for older adults in the context of Historytelling?") a workshop with three different groups was conducted.

Due to possibly low computer literacy among the participants, the technology was partly replaced by a real-life example. (see also Lindsay et al., 2012; Sengpiel et al., In Press)

There are a variety of methods using real life examples as prototypes for technology development, among them „invisible technology videos" (Lindsay et al., 2012), (Cultural) Probes (Brandt and Grunnet, 2000), and Forum Theater (Rice et al., 2007).

We used a simulation game often used in educational context, more specifically a modified simulation game used by Reich (2007).

He states that the ideal simulation game consists of seven phases: (1) introduction, (2) information and reading, (3) opinion-forming and strategy planning, (4) interaction within the groups, (5) preparation of a plenum, (6) conducting a plenum, (7) game evaluation. Due to a lack of time, the second and third phases were omitted during the workshop and conducted a priori by the researchers. Phase seven was conducted by the researchers after the workshop.

We recruited nine older women (M=68) through the "Deutscher Frauenring e.V.", a leading women's organization in Germany, who took part in three rounds within a larger full day workshop with several parts on the University campus.

To record interactions, we used a desktop microphone and the software "Speak a Message" running on a laptop with external screen and mouse.

The simulation game lasted 15 min per round plus seven minutes for discussion. Participants were interviewed afterwards according to their respective roles:

Assistant (Please simulate a voice assistant. Remain within your role and react to anything you notice.)

Storyteller (Please read out loud this shortened version of "Mother Hulda". The assistant will help you with the recording.)

Observer (Please observe the interaction between the assistant and the storyteller and fill in this observation sheet.)

The assistant and the storyteller were positioned to have no direct eye contact, while the observer was asked to sit seeing both (see the sketch and photo in figure 1).

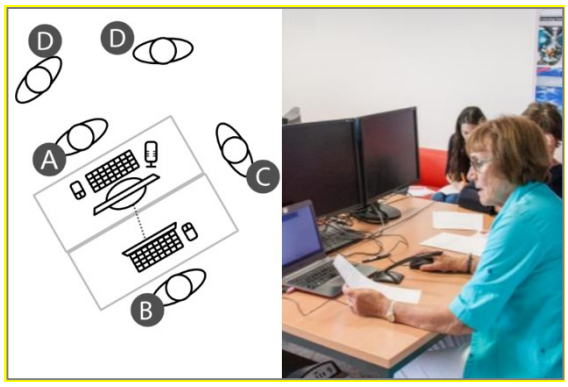

Figure 1: Sketch and photo of the simulation game's setup; A=Storyteller, B=Assistant, C=Observer, D=examiner.

3.2 Results

The use of the simulation game method showed that participants were good at taking the provided perspectives, were eager to give meaningful information and help with their expertise and had no problems solving the tasks given.

In the follow-ups there were lively discussions about possible improvements, which will be translated into requirements for the assistance system.

3.2.1 Participants

All the 9 older participants were women. Eight out of nine older adults were using their computer or laptop frequently, all participants own a smartphone and seven out of nine used it frequently. Technology is mostly used for communication and targeted information research. As expected for their age group, they scored relatively low on the computer literacy scale (CLS: M=16, SD: 3.67) but high on the affinity for technology scale (ATI: M= 3.8, SD=0.8).

3.2.2 Simulation Game

Simulation game results are quite diverse between groups, for they showed very different behavior. For example, group 2 had a fluent dialogue, while the other groups had rather functional dialogues, e.g.:

Group2: *Assistant: "I am the voice assistant. My name is…" Storyteller: "I am the storyteller. My name is… and I will start right away."*

Groups 1 & 3: *Assistant: "I am the voice assistant. My name is… Have you turned on your microphone?" Storyteller: "Yes, should I press the record button? Assistant: "Yes "*

Group 1 did not establish a fluent dialogue, yet in the interview the storyteller said she would have liked a more fluent dialogue and better feedback from the assistant, especially regarding recording quality.

Group 2 established a fluent dialogue from the start and immediately reacted to the assistant's remark to speak louder. However, in the interview the storyteller considered this interruption unpleasant and said she would prefer visual help and remarks, for any interruption in the flow of storytelling should be avoided.

Group 3 started with a longer dialogue, but the storyteller had forgotten to record it. The assistant said in the interview that she had noticed it, but did not want to interrupt the storyteller, conceding afterwards that it would have been better to do so. They also appreciated the dialogue in the beginning and wished it could have been continued in the study as well as with the technical system to be developed.

3.2.3 Resulting Interface Requirements

With the simulation game, some requirements were developed for the assistance system: It should answer user questions with a fluent verbal dialogue, being able to assess events' relevance and adapt kind and timing of communication to avoid unnecessary interruptions. In essence, the participants hoped for an assistance system behaving like a polite competent human, perhaps pushing the boundaries of today's technology.

Especially the recording flow should be supported from start to finish. There are further requirements for voice input communication in the literature, which were pointed out in 1.1

3.3 Resulting Interface

The resulting high-fidelity prototype is based on an interface presented in an already published paper (Volkmann et al., 2018) to ensure consistency within the HT project. Since our prototype could not display dynamic content, some interface elements had to remain static. Thus, some interactions such as providing feedback in recording sessions were triggered by the experimenter as Wizard of Oz. There were four kinds of feedback:

- A visualization based on a VU (volume units) meter which is a standard display for the signal level in audio equipment (see figure 2).
- Warning messages (see figure 3).
- An earcon (ear + icon) which are "abstract, synthetic and mostly musical tons or sound patterns that can be used in structured combination" (Dingler et al., 2008).

Figure 2: VU meter used for audio visualisation.

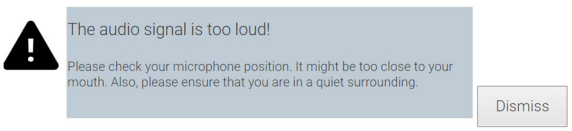

Figure 3: Warning message.

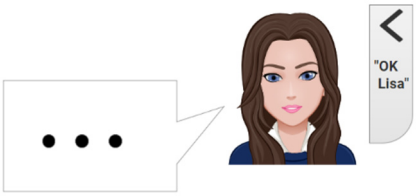

Figure 4: Voice assistant Lisa speaking.

- A voice assistant in form of an ECA as described in 1.2 (see figure 4).

The assistance is provided in three standardized, consecutive steps. First, a problem in audio quality is visualized through the VU meter. If the user does not perceive the problem and thus does not deal with it, an earcon is played and an additional warning message was displayed that the recording will be stopped. For the last step, the assistance varies. In a first implementation, there is no additional warning. In a second implementation, an avatar is used to give the information about the problem.

4 EVALUATION

To assess usability and user experience of the interface, a wizard of oz evaluation was conducted. The participants were given the task to record a story with the provided interface and assistance was provided as described in 3.3.

Two questions were essential for the evaluation: (1) Has the assistance been perceived? (2) Which assistance was preferred?

In the evaluation, eight older adults aged between 61 and 70 (M=66, SD=3) took part, five were male and three female. Six participants had already participated in the first study. They were recruited through personal contacts, mailing lists and notice boards. All evaluations took place at the university. To give the evaluation an attractive context, a Christmas flower and a candle were placed around the participants, among other things. Also, cake, water and hot drinks were served (Newell et al., 2007). Figure 5 shows part of the room in which the evaluation was conducted. Behind the participant (A), the wizard (B) and the recorder (C) were present in the room.

First, the participants were handed out a questionnaire regarding demographic information, affinity for technology (Franke et al., 2018) and computer literacy (Sengpiel and Dittberner, 2008). Then, the participants were confronted with the voice input interface. The order of assistance use was randomized. Before each run, the microphone was secretly placed too far apart from the participants creating a problem with audio quality, to justify the system warning and trigger a response from the participants. Before the run of the classic interface, also the microphone cable was unplugged. After the recorded interaction with the interface, the User Experience Questionnaire (UEQ) (Schrepp, 2015) was filled out by the participants. The second interface was tested correspondingly. In a post-interview possible adjustment possibilities and preferred interfaces were discussed.

4.1 Results

Overall, the Wizard of Oz prototype proved well suited to test the functionality that would have been hard to implement, such as the transcription or the avatar, although wizard response time was sometimes too high to satisfy the participants. The rather simple prototype allowed for participants' immersion in the process of storytelling.

4.1.1 Participants

Of the 8 older participants (61 – 70 years, M=66, SD=3), 5 were women and 3 were men. They mainly used computers, tablets and smart phones, mostly for text editing, email, internet searching and surfing. For their age group, they had relatively high computer literacy (CLS: M=21.8, SD: 4.1) and high affinity for technology (ATI: M= 3.1, SD=1.1).

4.1.2 Awareness of Provided Feedback

The participants used the provided VU meter for regular monitoring. The earcon was often ignored or not perceived at first, especially by the participants immersed in the storytelling. Three out of eight participants perceived the earcon from the start, three other participants perceived it the second time. It seems reasonable to assume that earcons need to be learnt before (Dingler et al., 2008).

All participants perceived the information the voice assistant gave about occurring problems, but they did not engage in conversation.

A combined approach considering the importance of intervention might work best in this scenario. It is generally difficult to give the storyteller information about options to improve the quality of the audio signal, while maintaining the oral fluency of the storytelling process.

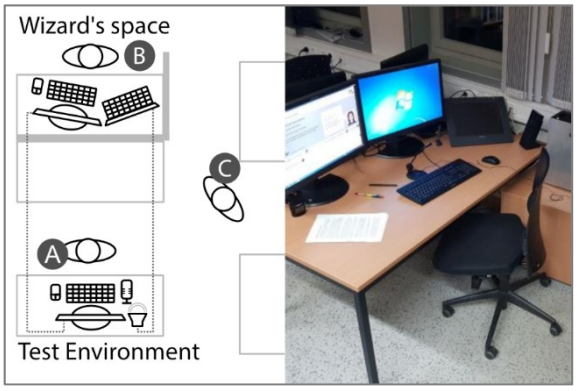

Figure 5: Sketch and photo of the evaluation's setup.

4.1.3 Preferred Assistance

The User Experience Questionnaire (UEQ) revealed a strong preference for voice assistance, but only if feedback was triggered on time. If it was delayed, then discomfort, confusion, and frustration occurred, and participants rated the User Experience much lower in all UEQ categories. However, using the interface without voice assistant, delays had much smaller impact on the UEQ score. Figure 6 shows this interaction effect for voice assistant x delay based on UEQ mean scores across the scales found in Table 2.

Table 2: Results of the User Experience Questionnaire (Scale ranging from -3 to +3) for recordings with and without voice assistant either delayed or on time, indicating an interaction effect (see figure 6).

	On time (N=5)		Delayed (N=3)	
Aspect	**M**	**SD**	**M**	**SD**
Recordings with voice assistant				
Attractiveness	1.37	1.1	-0.7	0.8
Perspicuity	1.5	0.8	-1.2	0.6
Efficiency	1.3	0.7	-0.3	0.3
Dependability	0.8	0.8	-1.2	0.4
Stimulation	1.15	0.5	0	0.8
Novelty	0.85	0.6	0.2	0.7
Recordings without voice assistant				
Attractiveness	0.77	1.0	0.3	1.0
Perspicuity	1.25	1.3	1.0	0.9
Efficiency	0.95	0.8	1.0	0.7
Dependability	0.35	1.3	0.5	1.1
Stimulation	0.95	1.3	0.3	1.2
Novelty	0.65	0.4	0.0	1.0

5 DISCUSSION

Following the HCD+ approach, possible future users were integrated in all steps of the development of the voice input component and methods were adjusted to accommodate user characteristics of older adults. Although the computer literacy score was rather high compared to other groups of older adults, due to the large heterogeneity within groups the adjustments were beneficial to the goal of universal usability, not to preclude anyone by design.

Regarding the first research question “Which type of voice input is suitable for older adults”, the subsequent transcription was preferred among the participants. Participants wanted to have both, text and recorded audio, which can be achieved by the transcription. However, the quality of the text to speech engine used in the tested software was not sufficient to maintain uninterrupted oral fluency and there were still errors in the transcript. Additional studies have to be conducted to assess the maximum acceptable fault tolerance and if current technology can undercut this line. If that is not possible with current technology, we suggest weighing the potential benefits to a loss in user experience due to user’s frustration.

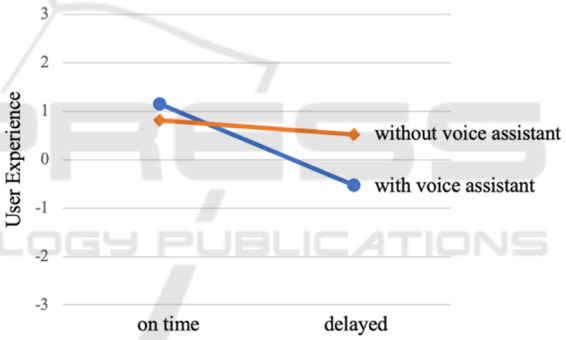

Figure 6: Interaction effect for voice assistant x delay on UEQ mean scores (see table 2 for details).

For practical purposes, the HT system could estimate in the beginning, whether recordings with voice assistant could be delivered without noticeable delay and conceal it otherwise to avoid the “UX penalty” for a delayed voice assistant shown in figure 6 and table 2. Audio files could be stored and transcribed later (with enhanced technology) as well. In the HT context, volunteers might also be willing to correct errors in the transcripts for the storytellers.

Regarding the second research question on minimal requirements for a conversational agent for older adults in the context of Historytelling, users should be guided through the recording process. A virtual speech assistant giving necessary information could be helpful, but it should recede into the background during story recording and graphical user interface elements should be used for regular

monitoring instead. Again, a delay in the assistants' feedback should be avoided, because it cripples user experience, inverting the benefits of conversational agents and leaving the users uncomfortable and confused.

ACKNOWLEDGEMENTS

We would like to thank Dr. Daniel Wessel for his help planning the evaluation and the many participants who sacrificed their spare time for the Historytelling project. The HCD+ approach would never work without them.

REFERENCES

Arik, S. O., Chrzanowski, M., Coates, A., Diamos, G., Gibiansky, A., Kang, Y., … others. (2017). Deep voice: Real-time neural text-to-speech. *ArXiv Preprint ArXiv:1702.07825*.

Bornat, J. (2001). Reminiscence and oral history: parallel universes or shared endeavour?. *Ageing and Society*, *21*(02), 219–241. https://doi.org/10.1017/S0144686X01008157

Brandt, E. and Grunnet, C. (2000). Evoking the future: Drama and props in user centered design. In *Proceedings of Participatory Design Conference (PDC 2000)* (pp. 11–20).

Cassell, J. (2000). Embodied conversational interface agents. *Communications of the ACM*, *43*(4), 70–78.

Cohen, P. R. and Oviatt, S. L. (1995). The role of voice input for human-machine communication. *Proceedings of the National Academy of Sciences*, *92*(22), 9921–9927.

Dingler, T., Lindsay, J. and Walker, B. N. (2008). Learnabiltiy of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. International Community for Auditory Display.

Eisma, R., Dickinson, A., Goodman, J., Syme, A., Tiwari, L. and Newell, A. F. (2004). Early user involvement in the development of information technology-related products for older people. *Universal Access in the Information Society*, *3*(2), 131–140. https://doi.org/10.1007/s10209-004-0092-z

Fisk, A. D. (Ed.). (2009). *Designing for older adults: principles and creative human factors approaches* (2nd ed). Boca Raton: CRC Press.

Franke, T., Attig, C. and Wessel, D. (2018). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction*, 1–12. https://doi.org/10.1080/10447318.2018.1456150

Hailpern, J., Karahalios, K., DeThorne, L. and Halle, J. (2010). Vocsyl: Visualizing syllable production for children with ASD and speech delays. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility* (pp. 297–298). ACM.

Hazen, T. J., Saenko, K., La, C. H. and Glass, J. R. (2004, October). A segment-based audio-visual speech recognizer: Data collection, development, and initial experiments. In *Proceedings of the 6th international conference on Multimodal interfaces* (pp. 235-242). ACM.

Isbell, R., Sobol, J., Lindauer, L. and Lowrance, A. (2004). The Effects of Storytelling and Story Reading on the Oral Language Complexity and Story Comprehension of Young Children. *Early Childhood Education Journal*, *32*(3), 157–163. https://doi.org/10.1023/B:ECEJ.0000048967.94189.a3

Isbister, K. and Doyle, P. (2002). Design and evaluation of embodied conversational agents: A proposed taxonomy. In *The first international joint conference on autonomous agents & multi-agent systems*.

Klein, L. (2016). *Design for Voice Interfaces*. USA, Sebaastopol, CA: O'Reilly.

Levin, G. and Lieberman, Z. (2004). In-situ speech visualization in real-time interactive installation and performance (p. 7). ACM Press. https://doi.org/10.1145/987657.987659

Lindsay, S., Jackson, D., Schofield, G. and Olivier, P. (2012). Engaging Older People Using Participatory Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1199–1208). New York, NY, USA: ACM. https://doi.org/10.1145/2207676.2208570

Muller, M. J. (2003). Participatory design: the third space in HCI. *Human-Computer Interaction: Development Process*, *4235*, 165–185.

Newell, A., Arnott, J., Carmichael, A. and Morgan, M. (2007). Methodologies for Involving Older Adults in the Design Process. In C. Stephanidis (Ed.), *Universal Acess in Human Computer Interaction. Coping with Diversity* (Vol. 4554, pp. 982–989). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73279-2_110

Orso, V., Spagnolli, A., Gamberini, L., Ibañez, F. and Fabregat, M. E. (2015). Involving Older Adults in Designing Interactive Technology: The Case of SeniorChannel. In *Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter* (pp. 102–109). New York, NY, USA: ACM. https://doi.org/10.1145/2808435.2808464

Ortiz, A., del Puy Carretero, M., Oyarzun, D., Yanguas, J. J., Buiza, C., Gonzalez, M. F. and Etxeberria, I. (2007). Elderly Users in Ambient Intelligence: Does an Avatar Improve the Interaction? In C. Stephanidis and M. Pieper (Eds.), *Universal Access in Ambient Intelligence Environments* (pp. 99–114). Berlin, Heidelberg: Springer Berlin Heidelberg.

Reich, K. (Hg.): Methodenpool. In: url: http://methodenpool.uni-koeln.de

Rice, M., Newell, A. and Morgan, M. (2007). Forum Theatre as a requirements gathering methodology in the design of a home telecommunication system for older

adults. *Behaviour & Information Technology*, *26*(4), 323–331. https://doi.org/10.1080/01449290601177045

Schafer, R. W. (1995). Scientific bases of human-machine communication by voice. *Proceedings of the National Academy of Sciences*, *92*(22), 9914–9920. https://doi.org/10.1073/pnas.92.22.9914

Schrepp, M. (2015). User Experience Questionnaire Handbook. *All You Need to Know to Apply the UEQ Successfully in Your Project*.

Sengpiel, M. and Dittberner, D. (2008). The computer literacy scale (CLS) for older adults-development and validation. In *Mensch & Computer* (pp. 7–16).

Sengpiel, M., Volkmann, T. and Jochems, N. (In Press). Considering older adults throughout the development process – The HCD+ approach. In *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2018 Annual Conference*. Berlin.

Sengpiel, M. and Wandke, H. (2010). Compensating the effects of age differences in computer literacy on the use of ticket vending machines through minimal video instruction. *Occupational Ergonomics*, *9*, 87–98.

Sutcliffe, A. (2017). Designing User Interfaces in Emotionally-Sensitive Applications. In R. Bernhaupt, G. Dalvi, A. Joshi, D. K. Balkrishan, J. O'Neill and M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2017* (pp. 404–422). Cham: Springer International Publishing.

Tsiourti, C., Joly, E., Wings, C., Moussa, M. B. and Wac, K. (2014). Virtual Assistive Companions for Older Adults: Qualitative Field Study and Design Implications. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare* (pp. 57–64). ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). https://doi.org/10.4108/icst.pervasive health.2014.254943

Tsiourti, C., Quintas, J., Ben-Moussa, M., Hanke, S., Nijdam, N. A. and Konstantas, D. (2018). The CaMeLi Framework—A Multimodal Virtual Companion for Older Adults. In Y. Bi, S. Kapoor and R. Bhatia (Eds.), *Intelligent Systems and Applications* (pp. 196–217). Cham: Springer International Publishing.

Volkmann, T., Dohse, F., Sengpiel, M. and Jochems, N. (2018). Age-Appropriate Design of an Input Component for the Historytelling Project. In *Congress of the International Ergonomics Association* (pp. 672–680). Springer.

Weiss, B., Wechsung, I., Kühnel, C. and Möller, S. (2015). Evaluating embodied conversational agents in multimodal interfaces. *Computational Cognitive Science*, *1*(1), 6. https://doi.org/10.1186/s40469-015-0006-9

Wood, L. E. (1997). Semi-structured Interviewing for User-centered Design. *Interactions*, *4*(2), 48–61. https://doi.org/10.1145/245129.245134.