

Corpus Analysis of Spoken Smart-Home Interactions with Older Users

Sebastian Möller¹, Florian Gödde¹ and Maria Wolters²

¹ Quality and Usability Lab, Deutsche Telekom Labs, TU Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany

² Centre for Speech Technology Research, University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW, Scotland

E-mail: sebastian.moeller@telekom.de, tsubasa@cs.tu-berlin.de, mwolters@inf.ed.ac.uk

Abstract

In this paper, we present the collection and analysis of a spoken dialogue corpus obtained from interactions of older and younger users with a smart-home system. Our aim is to identify the amount and the origin of linguistic differences in the way older and younger users address the system. In addition, we investigate changes in the users' linguistic behaviour after exposure to the system. The results show that the two user groups differ in their speaking style as well as their vocabulary. In contrast to younger users, who adapt their speaking style to the expected limitations of the system, older users tend to use a speaking style that is closer to human-human communication in terms of sentence complexity and politeness. However, older users are far less easy to stereotype than younger users.

1. Introduction

Spoken dialogue systems (SDSs) have reached a sufficient level of maturity for general use in specific domains, e.g. for travel information, telephone banking, or remote control of domestic devices. However, so far, SDSs have mainly been developed with healthy younger users in mind. Moreover, most existing corpora of human-machine interactions contain very little data from older users.

Older people differ from younger users in two important ways. Firstly, a person's learning, cognitive and/or sensory capabilities tend to decline with age (Rabbitt et al. 2005). Secondly, older people may well be less familiar and/or comfortable with computers and technology than younger users (Reynolds et al. 2002). Both factors potentially influence how older users approach a SDS in terms of vocabulary and language structure. Because systems are frequently developed on the basis of corpora obtained from younger users, the linguistic interaction behaviour of older users is frequently not in line with the inputs expected by the system. It is important to differentiate between the impact of cognitive abilities and attitude to technology on one hand and age on the other. It is well known that chronological age is not a very reliable predictor of physiological, anatomical cognitive function, since older users vary widely in their ability levels (Arking 2005, Rabbitt et al. 2005). If our systems adapt to users on the basis of age alone, we risk providing inadequate support to a 19-year-old with bad short-term memory, while unnecessarily restricting the 70-year-old professor emeritus of computer science who contributes to open source software in his spare time.

There is a small, but increasing amount of research on developing SDS for home-care and tele-care applications. Examples include scheduling appointments over the phone (Zajicek et al. 2004, Wolters et al., submitted), interactive reminder systems (Pollack, 2005), symptom management systems (Black et al. 2005) or environmental control systems (Clarke et al. 2005). Most studies focus on describing and analysing existing systems. However,

to date, very few studies have attempted to describe how older users communicate with spoken dialogue systems, and to compare older users' interaction style with that of younger users. This has been mainly due to a lack of transcribed corpora, which is beginning to be addressed (Cucchiari et al. 2006; Georgila et al. 2008).

In this paper, we present a corpus that was collected during a laboratory-based experiment in which both older and younger users interacted with a SDS for controlling domestic devices. Two research questions guided the design of the experiment:

- 1) Is there any difference between younger and older users in the way they talk to the SDS? If yes, are the observed differences mainly due to age, or are cognitive ability, affinity with technology, and/or experience with such technology more important?
- 2) To what extent are older users primed by the system, e.g. in terms of vocabulary usage? Can alignment towards the system's vocabulary be supported by appropriate help prompts?

We present a first attempt at answering the first question, highlighting some key differences in speaking style between younger and older users. With respect to the second question, we investigate whether there are changes between the first and second interaction. If exposure to help prompts helps shape users to adapt their speaking style to the kinds of input understood by the system, we should see significant changes in relevant linguistic variables between the two interactions regardless of help prompt type.

In Section 2, we present the set-up for collecting the corpus, including the smart-home system, the test participant pool and the experimental design. Section 3 describes how the data has been annotated, and Section 4 provides the analysis of the obtained corpus with respect to the general interaction behaviour, the speaking style and the adaptation to the system, for both user groups. We discuss our findings in the light of the two research questions in Section 5 and present some ideas for future work in Section 6.

2. Corpus collection

2.1 Environmental control SDS

We collected data with a Wizard-of-Oz environment (WOZ). The WOZ system was based on a full end-to-end SDS which covers a variety of use cases that relate to the control of objects and appliances in the home. The basis for this system was developed in the frame of the EU-funded IST-project INSPIRE (INfotainment management with SPeech Interaction via REmote microphones and telephone interfaces; IST 2001-32746); for a more detailed description, see Möller et al. (2004). The system allows speech-based control of lamps, blinds, heating/fan, the TV, a video recorder, an electronic program guide, and an answering machine. The users' speech is recorded by a microphone array or a portable wireless lapel microphone, recognized by a commercial speech recognizer, and interpreted by a key-expression matching module. As the speech recognizer has not been specifically trained on older users' speech, recognition performance for this user group was likely to be lower (Anderson et al. 1999). This increased need to compensate for speech recognition errors could potentially make the system more difficult to operate for older users, which would skew the results in favor of the younger user group. Therefore, we decided to replace the recognizer by a transcribing wizard during the experiment, providing equally close-to-perfect recognition performance for all user groups. The collected speech corpus will form a basis for speech recognizer adaptation in a later experiment.

The dialogue is managed via a set of generic dialogue nodes connected through a local and a global branching logic, as implemented in the rapid dialogue prototyping methodology of Rajman et al. (2004). Each of these nodes can play dialogue-specific help prompts, either each time a prompt is played (all-help option), or only when a misunderstanding has been detected by the system (limited-help option). Note that in the all-help condition, the system does not play a help prompt if the user initiates a new task by giving the correct command immediately; in that case, the system merely confirms execution. Initiative is mixed between user and system until a specific device and a type of action have been specified. If the task is complex, such as operating the answering machine, the system then takes the initiative to guide the user through further steps like TV program selection or message administration. System output is provided by concatenating pre-recorded messages uttered by a male speaker. In addition, longer lists of options (e.g. program listings) are displayed on the TV screen when necessary. The system is implemented in a room decorated as a living room at Deutsche Telekom Labs, Berlin.

2.2 Participants and test set-up

We collected data from 31 participants, 15 users aged 62 to 85 years (mean 68 years, referred to as "older users" in the following), and a comparative group of 16 participants between 22 and 29 years (mean 26 years, referred to as

"younger users"). An additional older participant took part in this experiment, but the interaction failed to be recorded.

The participants were first familiarized with the purpose of the experiment, the smart-home system and its capabilities. They then answered an initial questionnaire with items related to their personal background, their experience with speech technology and with domestic devices, as well as their general technical experience and attitude towards technology. A subsequent short story served as a kind of mind-setting for illustrating the usefulness of a speech-controlled smart-home system. The participants then carried out two scenario-guided interactions with the system. Each scenario consisted of several tasks addressing different devices and actions to perform, embedded in a "story" and balanced for complexity. After each task had been completed, users were asked to fill in a report card assessing their interaction with the system.

After each interaction, the participants had to fill out a questionnaire with 37 items related to their current experience with the system. The questionnaire was designed according to ITU-T Rec. P.851 (2003) and provided judgments on task effectiveness, speech understanding, interaction behavior of the system, dialogue symmetry and length, personal impression on the user, and system usability. After both interactions, a post-experimental questionnaire collected information on the overall impression and the fulfillment of expectations. Finally, an optional memory span test was carried out (forward and backward digit span tests) which assesses an important aspect of cognitive processing, the capacity of a short-term store for information processing. Audio and video recordings of all dialogues were collected for later annotation and analysis.

	Older	Younger	Sig.
Memory Span Score	12±2.6	19.4±2.9	p<0.0001
Technology Affinity	0.23± 0.5	0.51± 0.3	p<0.005
Experience w/ Technology	8	14	n/a

Table 1: Key properties of younger and older user groups.

10 older participants and 16 younger participants took the digit span test. The key properties of both user groups are summarized in Table 1. As expected, there is a highly significant difference in digit span between older and younger users. Older users have a smaller short-term memory capacity than younger users, which affects their ability to remember the information presented in a prompt. All but two younger participants had already experience with SDSs, but only half of the older participants. The average of the initial questionnaire's ratings on technical affinity showed that the older users show a significantly lower affinity to technology than the younger users).

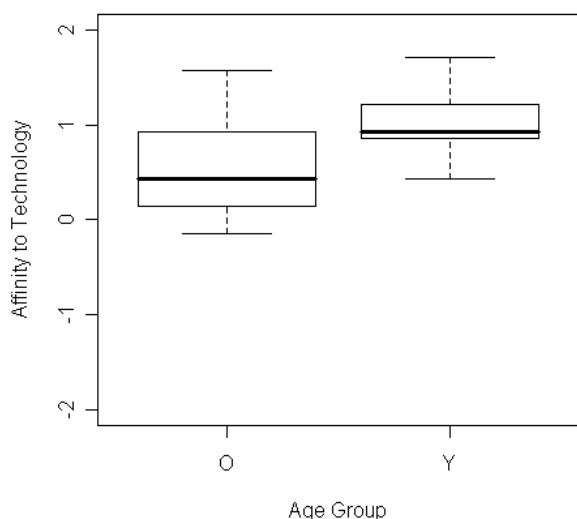


Figure 1: Affinity to technology per age group.

However, this finding should not be misinterpreted to mean that all older users have a low affinity to technology. The boxplot in Figure 1 shows that the main difference between older and younger users is that younger users have a higher affinity to technology in general, whereas the scores for older people may vary substantially. This illustrates that we cannot assume all older people to be technophobes. Instead, affinity to technology is affected by a complex web of cognitive and social psychological factors (Czaja et al. 2006).

3. Corpus annotation

All interactions were transcribed orthographically using the tool Transcriber (<http://trans.sourceforge.net>). The transcription conventions were based on the guidelines that were developed for the creation of the AMI meetings corpus (Moore et al. 2005; Carletta 2007), which has been used extensively for speech recognition research. Subdialogues corresponding to task boundaries are marked by Transcriber “report” tags. In each turn, keywords were marked as a “keyword” entity. We defined keywords as content words that could be interpreted as system commands. Task success and task failure were annotated by tagging the utterances that acknowledged success or signalled failure.

From these annotations, we derived a number of measures. The measures most relevant for the analyses reported here are:

- No. turns: Number of user turns.
- Shared Vocabulary (*SV*): Number of keywords shared by user and system utterances divided by the number of keywords uttered by the user only.
- Keyword-to-utterance-Length Ratio (*KLR*): Number of keywords divided by number of tokens in an utterance.
- Relative frequencies of groups of lexical items (Table 2)

Name	Description
art	Definite article
aux	Auxiliary verb form
dial	Dialectal form (e.g. “ick”=“ich”);
ich	Form of first person personal pronoun
intj	Hesitations and related particles
konj	Conjunctions
social	Markers of social interaction (e.g. please, thank you)
yes/no	Forms of yes and no

Table 2: Types of lexical items analysed.

4. Analysis

4.1 Interaction behaviour

Interaction behaviour was first analyzed on a global level, annotating the three usability dimensions effectiveness (here expressed in terms of task success), efficiency (dialogue duration), overall impression of the system, and overall rating of the interaction itself (both interaction questionnaires). There were significant differences between older and younger users both with regard to effectiveness and efficiency. Overall, older users produced more task failures and took longer to complete each interaction than the younger ones. Although the global impression after the experiment was more positive for the younger than for the older users (c.f. Table 3), the difference was not significant. However, when pooling the questionnaire items that relate to the quality of the spoken interaction with the system, younger users judged the interaction with the system significantly more positively than older users (c.f. Table 3).

	Older	Younger	Sig
No. failed tasks (overall)	3±2	1±1	0.05
No. errors (overall)	12 ±5	10± 5	n.s.
No. user turns (overall)	91± 21	76± 17	0.05
Overall impression	0.31±0.5	0.76±0.8	n.s.
Rating of Interaction	0.39±0.5	0.89±0.8	0.05

Table 3: Key differences in interaction behavior between younger and older users (mean ± std.dev.).

4.2 Speaking style

The analysis of speaking styles we present in this paper refers to results pooled by users. Since there are less than 20 data items in each age group, this level of granularity only allows us to detect very strong differences in speaking style with large effect sizes. Clear tendencies in the expected direction with medium effect sizes will not be significant. In the following tables, we will therefore give effect size (based on pooled variance, Cohen 1988) as well as significance.

	Older	Younger	Eff.Size	Sig.
No. Words	305±186	145±46	1.18	0.005
No. Words / Turn	3.4±1.7	2.0±0.4	1.08	0.05
<i>SV</i>	63.2±8	65.3±8	-0.21	n.s.
<i>KLR</i>	59.6± 18	78.6±15	-0.97	0.005

Table 4: Differences in speaking style between younger and older users (mean ± std.dev.).

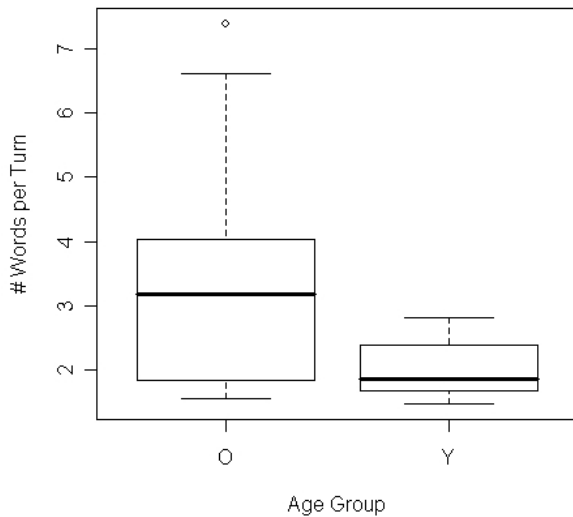


Figure 2: Number of words per turn for both user groups.

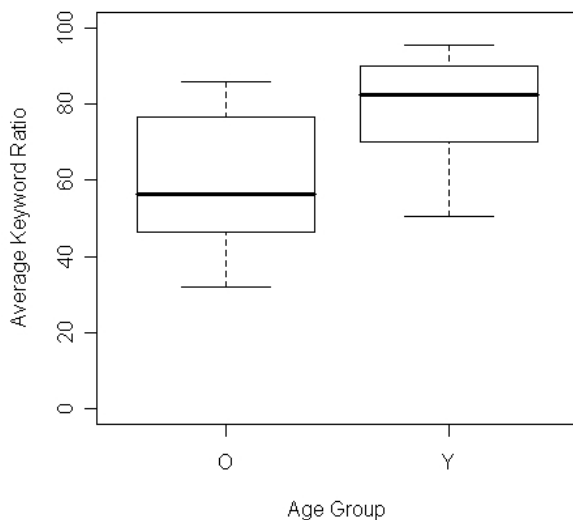


Figure 3: Average keyword ratio (KLR) for both user groups.

Older and younger users differed markedly in their speaking styles. Relevant overall differences are summarised in Table 4. Overall, younger users are more likely to use a “command style” language that is pared down to the vocabulary that the system is expected to understand. They use fewer words per turn and have a higher keyword ratio. However, these tendencies do not mean that older users are always verbose. As the boxplots

in Fig. 2 and Fig. 3 show, quite a few of the older users are just as concise as younger people. This fact is not necessarily observable from the standard deviations alone. Overall, the behaviour of younger users is more homogeneous and predictable than that of our older sample, as would be expected from the literature.

Older and younger users also differ significantly in the frequency of many of our linguistic markers. Table 5 shows the frequencies of these markers relative to the total number of words produced. Older people use more definite articles, more auxiliaries, more first person pronouns and, most importantly, more lexical items related to social interaction, such as “please” and “thank you”. Incidentally, there are no significant differences in the frequency of dialect markers. This is due to the fact that most of the dialectal forms in our corpus (38/41 or 93%) are contributed by three people, an older male (28/41; 68%) who has one of the lowest observed scores for affinity to technology, -0.14, another older male (6/41; 15%) and an older female (4/41; 9%). The other three word forms are produced by two older and one younger user. Thus, even though almost all dialect forms are produced by older users, their occurrence is highly idiosyncratic.

	Older	Younger	Eff. Size	Sig.
art	9.3±4.3	5.3±2.7	1.12	0.01
aux	3.3± 2.8	1.4±1.7	0.84	0.05
dial	0.5±1.2	0.4±0.1	0.61	n.s.
ich	2.8±2.3	1.4±2.1	0.61	0.05
intj	2.5±3.1	2.9±3.0	-0.14	n.s.
konj	2.3±1.6	1.3±1.3	0.65	n.s.
social	6.8±4.6	3.2±4.7	0.78	0.01
yes/no	5.1±3.0	8.1±4.7	-0.77	n.s.

Table 5: Differences in relative frequency of lexical items between younger and older users (mean ± std.dev.).

	Keyword Ratio	Sig.	Shared Voc.	Sig.
art	-0.79	0.0001	-0.56	0.005
aux	-0.63	0.0001	-0.51	0.005
dial	-0.34	n.s.	-0.06	n.s.
ich	-0.76	0.0001	-0.51	0.005
intj	-0.24	n.s.	-0.08	n.s.
konj	-0.71	0.0001	-0.53	0.005
social	-0.68	0.0001	-0.28	n.s.
yes/no	0.12	n.s.	-0.07	n.s.

Table 6: Correlations between lexical items and keyword ratio / shared vocabulary (Spearman’s Rho).

Four of our linguistic markers are significantly correlated with keyword ratio and shared vocabulary: definite articles, auxiliaries, frequency of first person pronouns and frequency of conjunctions. All four lexical categories correspond to parts of speech that are part of verbose

descriptions of what users want the system to do, but they can easily be left out of bare commands of the form noun plus infinitive. In addition, while a lower frequency of social interaction markers results in an improved keyword ratio, we found no correlation between the amount of shared vocabulary and users' tendency to use expressions more suitable for human-human interaction when talking to the INSPIRE system.

4.3 Adaptation to the system

Table 7 shows that users adapt their speaking style to suit the system's requirements in their second interaction regardless of the type of help prompt they received in the first interaction. Overall, older users tend to adapt more than younger users, in particular with respect to two groups of lexical markers that are associated with higher keyword ratio, the relative frequency of first person pronouns and the relative frequency of social interaction markers. Older users are far more likely to avoid those words in the second interaction than younger users.

Although the individual changes in speaking style exhibited by older users all point in the right direction, none of them is significant at $p < 0.05$ or better, although some changes (increase in shared vocabulary, decrease in number of articles, first person pronouns, and social interaction markers) are significant at the 0.1 level. Overall, younger users do not tend to adapt their speaking style significantly from the first interaction to the second. The only exception is the frequency of definite articles, which decreases significantly ($p < 0.05$).

	Older	Younger	Eff.	Sig.
<i>SV</i>	6.7±9.9 (.)	2.5±11.7 (ns)	0.38	n.s.
<i>KLR</i>	10.9±7.9 (ns)	6.7±8.7 (ns)	0.51	n.s.
art	-4.9±4.8 (.)	-2.0±4.1 (*)	-0.63	n.s.
ich	-2.2±2.0 (.)	-0.4±1.6 (ns)	-1.0	0.05
dial	-0.3±1.1 (ns)	-0.1±0.3 (ns)	-0.25	n.s.
social	-3.8±3.9 (.)	0.6±4.1 (ns)	-1.09	0.01

Table 7: Comparison of speaking style changes between scenarios for younger and older users (mean ± std.dev.).

Significances at a 0.05 level are marked with (*), at a 0.1 level with (.)

5. Discussion

Our results clearly indicate that older users use a different speaking style and a different vocabulary when addressing our speech-based smart-home system than younger ones. In contrast to younger users, who adapt their speaking style to the expected limitations of the system, older users tend to use a speaking style that is closer to human-human communication in terms of sentence complexity and politeness.

Even though these general differences are as expected, our results also show that older users are far less easy to stereotype than younger users. Even though standard deviations are overall fairly similar, the boxplots in

Figures 2 and 3 show that for the same variable, the overall range of older users' values is much wider than that of younger users.

Our analyses also suggest that older users can learn how to speak to a dialogue system if given context-sensitive help when errors are encountered. This confirms well-established results that were achieved with younger users (e.g. Zoltan-Ford 1991). We even found indications that older users adapt more readily than younger ones, whose behaviour essentially changes very little. There are several possible reasons for this effect: it could be that younger users already performed more or less near ceiling, and so did not need to adapt as much as older people, or it could be that we recruited an exceptionally highly experienced set of younger users who already knew how to interact with spoken dialogue systems.

6. Future work

We are currently working on a more detailed annotation and analysis of the audio, video, and transcripts of all 64 interactions recorded in our corpus.

The audio files collected from both user groups will be used for the adaptation and assessment of the speech recognizer which forms part of the final system set-up. We also plan to analyse the extent to which users adapt to the acoustic characteristics of the male INSPIRE voice, such as pitch range and mean fundamental frequency. Such accommodation is well-documented in both human-human and human-computer interaction (Oviatt et al. 2004).

To facilitate a richer analysis of users' linguistic interaction behaviour, we are planning a more detailed linguistic annotation of the transcripts with dialogue acts, parsing, and part-of-speech.

Finally, in order to investigate types of interaction problems encountered by both user groups (see Oulasvirta et al. 2006), the corpus will be annotated with error types following the scheme set out in (Möller et al. 2007). These interaction problems form the basis for novel user simulation techniques which will be used in an automatic usability evaluation workbench (Möller et al. 2006).

We plan to make all data available to the scientific community at the end of the MeMo project, when the data analysis and annotation has been completed.

7. Acknowledgements

This study was carried out within the project MeMo (Usability Workbench for Rapid Product Development) funded by Deutsche Telekom AG. The authors would like to thank the whole MeMo team for their support, in particular Marc Hümmer (Fraunhofer FIT, Birlinghofen) and Thimios Dimopoulos (formerly DAI-Labor, TU Berlin) for setting up and running the experiment. The MATCH project (Scottish Funding Council grant no. HR04016) funded Maria Wolters' contribution as well as a two-week research visit to Edinburgh by Florian Gödde.

8. References

- Anderson, S., Liberman, N., Bernstein, E., Foster, S., Cate, E., Levin, B., Hudson, R. (1999). Recognition of elderly speech and voice-driven document retrieval. In *Proc. ICASSP*, Phoenix, Arizona.
- Arking, R. (2005). *The Biology of Aging*. 3rd edition. Oxford, UK: Oxford University Press.
- Black, L.A., McMeel, C., McTear, M., Black, N., Harper, R., Lemon, M. (2005). Implementing autonomy in a diabetes management system. *J Telemed Telecare*, 11 Suppl 1:6–8.
- Carletta, J. (2007). Unleashing the killer corpus: Experiences in creating the multi-everything AMI Meeting Corpus. *Language Resources and Evaluation*, 41(2):181–190.
- Clarke, K., Lewin, M.R., Atkins, D., Kalawsky, R.S. (2005). Testing a framework for multimodal control in the home environment. In *Proc. Perspectives in Pervasive Computing*, pp. 87–95.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd edition. Hillsdale, N.J.: Lawrence Erlbaum
- Cucchiari, C., van Hamme, H., van Herwijnen, O., Smits, F. (2006). Jasmin-CGN: Extension of the spoken dutch corpus with speech of elderly people, children and non-natives in the human-machine interaction modality. In *Proc. 5th International Conference on Language Resources and Evaluation*, Genoa, Italy, pp. 135–138.
- Czaja, S.J., Charness, N., Fisk, A.D., Hertzog, C., Nair, S.N., Rogers, W.A., Sharit, J. (2006). Factors predicting the use of technology: Findings from the center for research and education on aging and technology enhancement (CREATE). *Psychology and Aging*, 21:333–352.
- ITU-T Rec. P.851 (2003). *Subjective Quality Evaluation of Telephone Services Based on Spoken Dialogue Systems*, International Telecommunication Union, Geneva.
- Georgila, K., Wolters, M., Karaiskos, V., Kronenthal, M., Logie, R., Mayo, N., Moore, J.D., Watson, M. (2008). A fully annotated corpus for studying the effect of cognitive ageing on users' interactions with spoken dialogue systems. In *Proc. 6th International Conference on Language Resources and Evaluation*, Marrakech, Morocco
- Moore, J., Kronenthal, M., Ashby, S. (2005). Guidelines for AMI speech transcriptions. *Technical report*, AMI Deliverable.
- Möller, S., Krebber, J., Raake, A., Smeele, P., Rajman, M., Melichar, M., Pallotta, V., Tsakou, G., Kladis, B., Vovos, A., Hoonhout, A., Schuchardt, D., Fakotakis, N., Ganchev, T., Potamitis, I. (2004). INSPIRE: Evaluation of a smart-home system for infotainment management and device control. In *Proc. 4th Int. Conf. on Language Resources and Evaluation (LREC 2004)*, Lisbon, Vol. 5, pp. 1603-1606.
- Möller, S., Englert, R., Engelbrecht, K., Hafner, V., Jameson, A., Oulasvirta, A., Raake, A., Reithinger, N. (2006). MeMo: Towards automatic usability evaluation of spoken dialogue services by user error simulations. In *Proc. 9th Int. Conf. on Spoken Language Processing (Interspeech 2006 – ICSLP)*, Pittsburgh PA, pp. 1786-1789.
- Möller, S., Engelbrecht, K.-P., Oulasvirta, A. (2007). Analysis of communication failures for spoken dialogue systems. In *Proc. 8th Ann. Conf. of the International Speech Communication Association (Interspeech 2007)*, Antwerp, pp. 134-137.
- Oulasvirta, A., Möller, S., Engelbrecht, K., Jameson, A. (2006). The relationship of user errors to perceived usability of a spoken dialogue system. In *Proc. 2nd ISCA/DEGA Tutorial and Research Workshop on Perceptual Quality of Systems*, Berlin, pp. 61-67.
- Oviatt, S., Darves, C., Coulston, R. (2004). Toward adaptive conversational interfaces: Modeling speech convergence with animated personas. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 11(3).
- Pollack, M.E. (2005). Intelligent technology for an aging population: The use of AI to assist elders with cognitive impairment, *AI Magazine*, 26(2):9-24.
- Rabbitt, P., Anderson, M.M., Bialystok, E., Craik, F.I. (2005). The lacunae of loss? Aging and the differentiation of human abilities. In *Lifespan Cognition: Mechanisms of Change*, chapter 23, F. Craik and E. Bialystok, eds., Oxford University Press, New York NY.
- Rajman, M., Bui, T.H., Rajman, A., Seydoux, F., Trutnev, A., Quarteroni, S. (2004). Assessing the usability of a dialogue management system designed in the framework of a rapid dialogue prototyping methodology, *Acta Acustica united with Acustica*, 90:1096-1111.
- Reynolds, C., Czaja, S.J. et al. (2002). Age and perceptions of usability on telephone menu systems. In *Human Factors and Ergonomics Society Conference*, 30 Sept.-4 Oct. 2002, Baltimore, MD, Human Factors and Ergonomics Soc.
- Wolters, M., Georgila, K., Logie, R., MacPherson, S., Moore, J.D., Watson, M. (2007). Making spoken dialogue systems accessible: It's a matter of choice, submitted.
- Zajicek, M., Wales, R., Lee, A. (2004). Speech interaction for older adults. *Universal Access in the Information Society*. 3(2):122-130
- Zoltan-Ford, E. (1991). How to get people to say and type what computers can understand. *International Journal of Man-Machine Studies*, 34:527–547.