Spoken Language, Conversational Assistive Systems for People with Cognitive Impairments? – Yes, If.

Ramin Yaghoubzadeh
CITEC, Bielefeld University
33501 Bielefeld, Germany
ryaghoubzadeh@uni-bielefeld.de

Stefan Kopp
CITEC, Bielefeld University
33501 Bielefeld, Germany
skopp@techfak.uni-bielefeld.de

ABSTRACT

We analyzed autonomous conversational spoken interaction as a modality for assistive systems in initial groups of older adults and people with cognitive impairments; we had previously explored this in a WOz setup. Solving a simple task in the domain of week planning, subjects readily interacted with the system. Performance was generally good, but dependent on successful adherence to a specific terse conversational style. Enforcing these patterns in a socially acceptable way during conversation is the next goal for the system.

Categories and Subject Descriptors

K.4.2 [**Computers and Society**]: Social Issues – assistive technologies for persons with disabilities

Keywords

Spoken interaction, cognitive impairment, daily assistant

1. INTRODUCTION

Both older adults and people with cognitive impairments can experience a higher quality of life when leading a self-determined lifestyle in their own home environment, but both groups may require assistance in some of the daily tasks that this entails. There are technological solutions for some areas, such as activity planning, home automation, facilitation of communication, or motivation for exercise [2]. Most of those solutions rely on textual information or the processing of task hierarchies and abstract visual symbols, which need to be acquired and can exclude a subset of the potential user groups.

In fact, many people are accustomed to requesting support using conversation with care personnel. This, together with the considerable advances of speech recognition in recent years, to a degree that acoustic models can cope with mildly dysarthric articulation better than ever before, opens the way for speech-controlled interfaces for wider user groups.

ASSETS '15, October 26–28, 2015, Lisbon, Portugal
ACM 978-1-4503-3400-6/15/10.
http://dx.doi.org/10.1145/2700648.281137.

Yet, current spoken interfaces generally either take the approach of navigating a constrained task graph step by step (such as the iconic pizza ordering systems), or of gathering all the information provided spontaneously and providing a contiguous chunk representing the most likely interpretation, after some processing (such as dictation systems or the systems familiar from modern phones). While the former provide low information density that can be suitable for people with cognitive impairments, they usually offer no leeway at all to escape the (often implicit) rigid patterns. Conversely, the latter approach accepts more unconstrained input, but does not guarantee appropriate information density. We study whether conversational approaches can combine the benefits of both, to a degree enabling their use by cognitively impaired people: the ability to offer a basic task structure to follow, to realize continous incremental interpretation of less constrained input, and to provide sufficiently transparent feedback, clarification requests or repairs.

2. STUDIES

2.1 Exploratory studies

In previous work, we explored the paradigm of a spoken conversational interaction with an embodied agent with both senior citizens and people with cognitive impairments (approx. ICD-10 F70), using a Wizard-of-Oz (covertly human-controlled) system in a week planning domain [3]. We found that the paradigm was very well received by the users and that they were generally willing and able to solve the given task of entering fictional appointments into a graphical calendar. We systematically introduced errors in the system replies, and employed two different sets of presentation and confirmation strategies to elucidate the limits for successful interaction, i.e. detection and repair of those errors. We found that the minimal information density (one piece of information per utterance, plus confirmation) was the most robust approach especially for the cognitively impaired users (86% repair rate as opposed to 50%) but was not rated less favorably than a more compact approach. In a control group of adults from a university environment, there were no differences between conditions since all errors were detected.

2.2 Conception of conversational system

The insights from the WOz experiments and subsequent conversational analyses led to the development of the flexdiam dialogue manager [4] that aims to fulfill all requirements for task-related conversational interfaces that are suitable also for people with cognitive impairments. The system

Figure 1: Setup and older participant (anonymized).

is able to yield at any time to the users' initiative, is aware of and able to cope with ambiguities resulting from misarticulation or users contradicting themselves, supports coping strategies to resolve repeated miscommunications and all basic forms of repairs employed in the task domains. Moreover, it supports introspection into the task structure, i.e. the system knows and can recapitulate where it is in the common task. Concurrent input and output includes speech but may extend to other modalities (currently, control of an embodied agent, a graphical display, and touch input).

2.3 Studies with autonomous system

In ongoing work, we run studies with this autonomous system. So far, we have recorded six senior participants (n=6, ages 78–87) [4] without noticeable cognitive impairments, and four people with cognitive impairments (n=4), insufficient for robust statistics but sufficient for valuable qualitative insights. We will extend the studies to another group of older adults and a different group of cognitively impaired people (n=12) who participated in the earlier experiments.

The system setup consisted of a speech recognizer (a locally running version of a leading brand), speech synthesis (CereVoice [1]), the dialogue system, and a screen with the agent and a graphical calendar (see Fig. 1). In the experiment, as in the earlier WOz experiment, participants solved the task of entering made-up appointments into a calendar using natural speech. They were provided with a sheet with iconic and textual representations of possible topics, to which they needed not adhere, but which were treated preferentially during speech recognition. Prior to the experiment proper, we conducted a short speech recognizer training session (<2 minutes of material), which completed successfully even for mildly dysarthric speakers. Participants were instructed to stay focused on the task, and that the agent would be able to only handle this topic. Interaction was split into two parts of up to three and at least five expected entries, respectively. Between phases, we scheduled an intervention where we could instruct users further if they could not communicate successfully with the system in the first phase. Instructions given here were related to refraining from paraverbal feedback ('u-huh'), between which the speech recognizer could not discern, and from verbose elaborations when asked for a specific bit of information.

All participants engaged readily in conversation with the system. In general, they were able to enter their fictional appointments [4]. Sometimes, concessions were made as to the wording of the entry. For one participant from each group, the system did not provide a pleasant and effective interaction. This was a direct result of the verbose interaction styles some participants employed; the interventions in the break between phases only made them change it for a short time before reverting to their previous style. The system in this iteration was configured to always yield the initiative to the human interlocutor.

3. DISCUSSION AND CONCLUSION

Conversational spoken interaction worked well unless the amount of information provided by the users overwhelmed the speech recognition and NLU modules. Always yielding the floor to the user only worked well for people that did not exhibit such behavior. Since explanatory interventions had no lasting effect, clearly a better solution would be to control the conversational structure in place. Naive strategies for this (such as simple heuristics over phrase lengths plus hard interruptions by barge-ins) would be efficacious, but would also violate the principles of cooperative and polite interlocutors and hence most likely harm long-term acceptance. Here, the employment of an embodied agent provides means of more subtle control of the pace, e.g. using gaze or gestural output, resorting to barge-ins only as a last resort. The design and selection of these strategies are the next areas of research carried out in a follow-up project, enabling a wider audience for conversational interfaces.

In summary, the results from the initial evaluations lend support to our assumption that spoken interaction can be a feasible interaction modality in assistive systems for persons with light to moderate cognitive impairment, using current speech recognition technology, and that the interactive, conversational approach is a suitable mechanism to prevent and repair miscommunications. Current work is focusing on socially acceptable conversational strategies for the system to control the flow of conversation to subtly increase adherence to structures the system can cope with.

4. ACKNOWLEDGMENTS

This research was partially supported by the German Federal Ministry of Education and Research (BMBF), and the Deutsche Forschungsgemeinschaft (DFG) in the Center of Excellence 'Cognitive Interaction Technology' (CITEC).

5. REFERENCES

[1] M. P. Aylett and C. J. Pidcock. *The CereVoice Characterful Speech Synthesiser SDK*. 2007.

[2] T. W. Bickmore, R. A. Silliman, K. Nelson, D. M. Cheng, M. Winter, L. Henault, and M. K. Paasche-Orlow. A randomized controlled trial of an automated exercise coach for older adults. *J American Geriatrics Society*, 61(10):1676–1683, 2013.

[3] R. Yaghoubzadeh, M. Kramer, K. Pitsch, and S. Kopp. Virtual agents as daily assistants for elderly or cognitively impaired people. In R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira, editors, *IVA 2013*, volume 8108 of *LNAI*, pages 79–91. Springer, 2013.

[4] R. Yaghoubzadeh, K. Pitsch, and S. Kopp. Adaptive grounding and dialogue management for autonomous conversational assistants for elderly users. In W.-P. Brinkman, J. Broekens, and D. Heylen, editors, *Proc. 15th Int Conf on Intelligent Virtual Agents, IVA 2015*, volume 9238 of *LNAI*, pages 28–38. Springer, 2015.