Designing Natural Language User Interfaces with Elderly Users

Stephan Schlögl
MCI Management Center
Innsbruck
Innsbruck, Austria
schlogl@mci.edu

Markus Garschall
CURE Center for Usability
Research and Engineering
Vienna, Austria
garschall@cure.at

Manfred Tscheligi
CURE Center for Usability
Research and Engineering
Vienna, Austria
tscheligi@cure.at

ABSTRACT

Speech and other forms of natural language based applications are increasingly accepted. In particular in situations where traditional interaction techniques, such as mouse and keyboard/keypad input, seem limiting or simply distracting, speech can offer an adequate alternative. A concrete application for this kind of technology can be found in Ambient Assisted Living solutions, which aim at supporting elderly people in their everyday activities. In this paper we want to highlight some of the challenges involved in designing speech technologies for this user group, and discuss areas where we see human factors research as being needed.

Author Keywords

Natural Language User Interfaces; Wizard of Oz;

ACM Classification Keywords

H.5.2 User Interfaces: Natural language; D.5.2 User Interfaces: Voice I/O

INTRODUCTION

Combining novel speech and language technologies into systems across diverse domains poses significant software engineering challenges. Assuring the usability and appropriate user experience of the resulting applications is an important and often overseen issue in this area. The use of Language Technology Components (LTC), i.e. for example Automatic Speech Recognition (ASR) and Text-to-Speech Synthesis (TTS), has significantly increased in recent years as their performance has improved. They have been integrated into several applications, with Interactive Voice Response (IVR) systems receiving widespread commercial deployment. Systems like Lets Go [15] and the DARPA communicator [21] were used to provide customers with schedule information over the telephone for flights [9], trains [11] and buses [19]. Furthermore, speech technologies have been integrated with Machine Translation (MT) in various prototypes that support multilingual communication during meetings [20], doctor-patient consultations [16] or while traveling [14]. LTCs are also used in settings in which traditional input and output modalities are less appropriate. Hands-busy eyes-busy situations, for example, require special interaction techniques. Hence, in-car navigation, route planning and other services such as electronic car manuals and interactive hotel reservation systems are now increasingly accessible via speech input and combined speech/map output [5]. Moreover, in the areas of tutoring systems and edutainment as well as in the health care sector [10] language technologies are used in several multi-modal/multi-lingual settings. Here mainly implementations of Embodied Conversational Agents show interesting application possibilities, including areas such as computer-based mental health care [2] and e-learning scenarios that help people learn about computer literacy or physics [7].

One driver of adoption to this new way of interaction can be found in increasingly ubiquitous access to products and services outside traditional office environments, where in many cases language technology solutions offer distinct advantages. Another contributing factor is the improved performance of these technologies, which has opened up new application areas in a variety of different fields. This trend is visible both from an application perspective, in the widespread use of voice dialing, in-car navigation systems with speech interfaces, instant web-based machine translation from mobile devices, and transactions accessed through IVR systems, as well as from a research perspective, in emerging areas such as Speech-to-Speech translation [18] and human-avatar interaction [1].

Yet, the design and development of speech based applications is still very technology driven. While in the early days of language technology human factors researchers were exploring the possibilities of natural user interfaces, their interest decayed quickly due to the small technological progress being made at the time. Only recently, with products such as Apple's Siri[1] and Google's Now[2], the interaction design community slowly gains back some interest. Still, there are many out there who see little future for speech and natural language based interaction. This is mainly due to the complexity involved in designing and building such technologies and the error proneness that often remains. Using the example of Ambient Assisted Living (AAL) this paper wants to highlight how HCI methods and techniques may be used to explore and advance current state-of-the-art language technology applications.

[1] http://www.apple.com/ios/siri/
[2] http://www.google.com/landing/now/

Workshop on Designing Speech and Language Interactions at CHI'14

SPEECH OPERATED ASSISTIVE TECHNOLOGY

In 2010, Europe alone counted 87 million people older than 65 years of age (i.e. EU-27) and projections see this number constantly increasing [3]. While significant improvements in the medical sector as well as an increased health awareness (e.g. regular sport, balanced nutrition, reduction of alcohol and tobacco) may generally assure our future well-being, rising age sooner or later does lead to physical and/or cognitive constraints. AAL solutions want to accompany this ageing process and support seniors in their everyday activities. The goal of AAL is to offer integrated technologies that can help people cope with their potentially deceasing physical and cognitive capabilities, such as smart living environments (e.g. BAALL[3]) or fall prevention solutions (e.g. iStoppFalls[4]). An important characteristic of these solutions is their increased use of speech and language technologies. Here speech is not only seen as a natural and straight forward interaction modality, but also as a possible solution to tackle the prevalence of fine-motor problems and/or chronic diseases. Statistics show that currently 1.2 million senior citizens in Europe suffer from Parkinson's[5] and 630.000 from multiple sclerosis[6]. The use of Natural Language User Interfaces (NLUI) allows for compensating these age-related restrictions. However, the construction of AAL speech interfaces requires both accurately working technology components and an interaction design that is well adapted to the distinct capabilities of the given user group.

DESIGNING NATURAL LANGUAGE USER INTERFACES

While it was already highlighted above that the potential employment of language technologies is undoubtedly increasing, several barriers remain to its widespread use. Language technology components are unavoidably imperfect and typically substantial engineering effort (gathering of corpora, training, tuning) is needed before prototypes involving such technologies can deliver a user experience robust enough to allow potential applications to be evaluated with real users. This is particularly true in the case of ASR where state-of-the-art recognition algorithms still show average word error rates (WER) of 20% [8] (10% within more specified domains [17]). Different accents and dialects may lead to even worse results [13].

User-centered design activities that focus on context information and domain specific interaction data could, however, be used to improve these error rates. For example, Fügen et al. [4] analyzed previously used material like transcribed speech from meetings and presentations to decrease the WER for public speeches and lectures. A similar approach can be applied to increase the quality of MT. Thus the first challenge where human factors research may help develop better, more accurate, language technology applications, is the collection and analysis of domain-specific interaction data (i.e. realistic corpora).

[3]http://www.baall.net

[4]http://www.istoppfalls.eu

[5]http://www.parkinsonsawareness.eu.com/campaign-literature/prevalence-of-parkinsons-disease/

[6]http://www.who.int/mental_health/neurology/Atlas_MS_WEB.pdf

In addition to having more and better annotated corpus data, some studies have shown that even if the used technology is flawed, people may be able to successfully accomplish given tasks. It has been demonstrated, for example, that for cases in which the ASR is poor, the overall user satisfaction can be increased through adapting the dialogue [12]. This highlights the second critical aspect where HCI can help building language technology applications i.e. appropriate dialog design. For both corpora collection and dialog design we require, however, adequate tools and methods. While for Graphical User Interfaces (GUI), methods like sketching and wire-framing allow us to test for varying effects and user satisfaction, and adapt the design based on the results, these low-fidelity prototyping techniques do not map well onto systems based around speech and other forms of natural language. Hence, applications that use language technology components such as ASR and TTS as their predominant interaction channel require a different design approach.

One technique that has been used to design language technology applications as well as to collect corpora is Wizard of Oz (WOZ). In general WOZ constitutes a prototyping method that is used by researchers and designers to obtain feedback on features that would require significant resources to be implemented. In a so called 'WOZ experiment' a human 'wizard' mimics the functions of a system, either entirely or in part, which makes it possible to evaluate potential user experiences and interaction strategies without the need for building a fully functional product first [6]. The application area of WOZ prototyping ranges from very low-fidelity, where researchers might refer to paper prototyping as being a form of WOZ, to high-scaled simulations in which the wizard only takes over a very distinct functionality of an envisioned product. Hence, in a way WOZ experimentation can be used to elicit user feedback throughout the entire development cycle of a new (language-based) system. It helps at the beginning by expanding/replacing low-fidelity prototyping methods such as interface sketches and storyboards, throughout the development process by simulating missing functionality and offering a way to compare different design solutions, as well as towards the end of a development cycle when it can be used to augment system components whose quality levels are not sufficient and therefore would require additional algorithm training and fine tuning. In addition to these three application levels, which focus mainly on the development and improvement of a system, WOZ can also be used as a tool to collect various types of domain specific corpus data. Hence, we have used it as an initial design method in a research project that aims at developing speech-based services for elderly people operating AAL services.

THE VASSIST PROJECT

The goal of the vAssist project[7] is to offer voice controlled home care and communication services for seniors based on a multi-lingual (i.e. French, German, Italian, English) NLUI. A particular focus lies on people who suffer from chronic

[7]http://vassist.cure.at

diseases and motor skills restrictions. The project is supported by the European Ambient Assisted Living Joint Program[8] and consists of 10 consortium members from industry and academia, located in three different countries (i.e. France, Austria, Italy). While vAssist predominately focuses on speech as an interaction modality it also seeks to optimize the integration of GUIs in situations where speech is not feasible or simply not wanted. In doing so it tries to lower potential adoption costs by using users' already existing hardware platforms (e.g. PCs, Smart TVs, mobile phones, tablets). A crucial aspect of vAssist is its multi-lingual/mutli-cultural perspective. With project partners from three different countries speaking three different languages the goal is not only to build the necessary (language) technologies but also to understand potential cultural differences in using them. Focus groups and user studies in all participating countries are used to gather continuous feedback and help identify areas of improvement. A user-centered, market-oriented design approach is followed, which assures that users are involved in all stages of design and development, leading to services and business models that are adapted to people's requirements and specific needs.

The following section discusses the first of two lab trials that were conducted as part of the vAssist project. Results of previous evaluations (e.g. focus groups) or technical aspects of the vAssist platform can be found in other publications listed on the vAssist website[9].

A LAB TRIAL WITH SENIORS

vAssist has planned several lab trials to be carried out throughout the design and development phase of the project. The first evaluations were conducted in October 2013. They were held in two different labs, one in Paris (France) and one Vienna (Austria), with both sites having their separate evaluation team. A mix of qualitative and quantitative methods was used to gather functional feedback and measure usability and user experience of two voice controlled software applications. In addition, these lab evaluations were used to collect natural speech data and to evaluate dialog designs. The focus lay on two tele-medicine services, which were designed and developed based on input from previously conducted focus group discussions. The first of the two services, i.e. the PILLBOX app, represents a voice controlled mobile phone application that allows for planning and recording medicine intake. It offers a reminder function and lets users report potential side effects. The second service, i.e. the DAILYCARE app, also runs on a mobile phone and serves as a diary for daily activities. It lets users register physical activities as well as pain and nutrition data. For the evaluations both apps were shown running on a computer screen. The GUI of the phone was mimicked using the image of a modern smartphone. Voice recognition and language understanding were simulated using the WEBWOZ Wizard of Oz Prototyping Platform[10], which was integrated with a JULIUS[11] ASR engine and a MARY TTS server[12]. Evaluations took place in a lab. Participants were told they would be interacting with a prototypical installation of a new system. The wizard was in the same room, however, participants were not told about the simulation until after the test. The recruitment of people was organized by end user organizations. In Vienna 7 elderly participants with fine motor-skill restrictions took part. In Paris we had 8 participants who were predominately suffering from chronic diseases. Eleven of these 15 participants were female and they were all between 60 and 86 years old (MEAN=71.8). For the evaluation they were asked to perform a predefined set of tasks using the speech user interfaces of our simulated tele-medicine services (e.g. changing a prescription, adding a reminder for a prescription, completing a report on sleep quality, etc.). In addition to observing the participants we used post-task/post-study interviews as a means for reflecting on usability issues and interaction aspects.

Our preliminary results show that elderly people generally like the idea of speech-based interaction. In fact, they find speech-controlled tele-medicine services a promising concept for solving the problem of motor-skill restrictions. They furthermore consider speech-based interaction as being less complicated when compared to more traditional forms of interaction, such as using a mouse or a keyboard. Also touch screens, as they are found in most of the currently available smartphones, in tablets and increasingly also in laptop and desktop computers, seem less intuitive to operate than using voice control. In particular, the potential of speech to provide a unified way of interacting with different services and different devices, was highly appreciated. Despite this overall positive attitude toward speech based interaction, we also observed a number of recurring usability issues, which kept participants from successfully completing some of the tasks. One reason for these problems was the particular way in which elderly people may talk. Long breaks between individual words (i.e. longer than usually observed with adults) caused the system to stop the speech recognition function before a user was able to complete a request or adequately answer a question. In addition, the strong local dialect of some of the participants led to a higher number of recognition errors. While with our WOZ setup we were able to mask these issues, it seems important to understand that for seniors technology components, in particular ASR, have to be specifically tuned and configured to achieve an acceptable performance.

Another aspect we found relevant is the initiation mechanism for the speech recognition function. As it is the case with most currently available voice operated consumer products, we used a button (i.e. a click button displayed next to the smartphone image) to open/close the microphone channel and consequently start/stop the speech recognition. Here, our observations revealed that most participants either forgot to press the button or did not release it before talking to the system. While they stated that they would still prefer a physical or a touch button over voice triggers or gestures for activating speech recognition, this highlights the need for clear visual or auditory feedback to indicate whether speech recognition

[8]http://www.aal-europe.eu

[9]http://vassist.cure.at/project_results/scientific_publications/

[10]https://github.com/stephanschloegl/WebWOZ

[11]http://julius.sourceforge.jp/en_index.php

[12]http://mary.dfki.de/

is on or off. Finally, during the post-study interviews, participants emphasized their demand for combined speech and touch interaction, i.e. mirroring the questions and answers of the speech-based interface on the smartphone or tablet device so that touch interaction may be available in cases where speech input fails.

SUMMARY AND FUTURE WORK

In summary our first insights into building natural language user interfaces for elderly people show that they favor speech over traditional physical input modalities. We found that speech based interaction leads to an increased engagement with assistive technology and therefore has the potential to improve the accessibility and technology acceptance of respective services. The employed evaluation method, which used WOZ in combination with existing language technology components, has proven successful. We could simulate a better system than what is currently available while at the same time being able to log and consequently analyses potential real-time problems.

Another round of lab trials as well as a long term field trial with the vAssist services are already in preparation and we hope that by the time of this workshop we will be able to present additional results. Furthermore, we would like to inform the community about our experience with involving end users in the design and development process. We believe that this will provide valuable input to discussions on how to improve user-centered development of natural language based computing systems.

ACKNOWLEDGEMENTS

The research presented in this paper is conducted as part of the vAssist project (AAL-2010-3-106), which is partially funded by the European Ambient Assisted Living Joint Programme and the National Funding Agencies from Austria, France and Italy.

REFERENCES

1. Bradley, J., Benyon, D., Mival, O., and Webb, N. Wizard of oz experiments and companion dialogues. In *Proceedings of BCS-HCI* (2010), 117–123.
2. Coyle, D., Doherty, G., Matthews, M., and Sharry, J. Computers in talk-based mental health interventions. *Interacting with Computers 19* (2007), 545–562.
3. European Union. Active ageing and solidarity between generations. In *Eurostat, Statistical books* (2011).
4. Fügen, C., Kolss, M., Bernreuther, D., Paulik, M., Stcker, S., Vogel, S., and Waibel, A. Open domain speech recognition and translation: lecture notes & speeches. In *Proceedings of ICASSP* (2000).
5. Geutner, P., Steffens, F., and Manstetten, D. Design of the VICO spoken dialogue system: Evaluation of user expectations by wizard-of-oz experiments. In *Proceedings of LREC* (2002).
6. Gould, J. D., Conti, J., and Hovanyecz, T. Composing letters with a simulated listening typewriter. *Communications of the ACM 26*, 4 (1983), 295–308.
7. Graesser, A. C., VanLehn, K., Rose, C. P., Jordan, P. W., and Harter, D. Intelligent tutoring systems with conversational dialogue. *AI Magazine 22*, 4 (2001), 39–51.
8. Jurafsky, D., and Martin, J. H. *Speech and Language Processing*, second ed. Prentice Hall, 2008.
9. Karpov, A., Ronzhin, A., and Leontyeva, A. A semi-automatic wizard of oz technique for Let'sFly spoken dialogue system. In *Lecture Notes in Computer Science: Text, Speech and Dialogue*. Springer, 2008, 585–592.
10. Keskin, C., Balci, K., Aran, O., Sankur, B., and Akarun, L. A multimodal 3D healthcare communication system. In *3D Conference* (2007).
11. Lamel, L., Rosset, S., Gauvain, J. L., and Bennacef, S. The LIMSI arise system for train travel information. In *Proceedings of ICASSP* (1999), 501–504.
12. Litman, D. J., and Pan, S. Predicting and adapting to poor speech recognition in a spoken dialogue system. In *Proceedings of AAAI* (2000), 100–111.
13. Liu, A. L., Hile, H., Kautz, H., Borriello, G., Brown, P. A., Harniss, M., and Johnson, K. Indoor wayfinding: Developing a functional interface for individuals with cognitive impairments. In *Proceedings of ASSETS* (2006), 95–102.
14. Paul, M. Overview of the iwslt 2009 evaluation campaign. In *Proceedings of IWSLT* (2009).
15. Raux, A., Bohus, D., Langner, B., Black, A. W., and Eskenazi, M. Doing research on a deployed spoken dialogue system: One year of Lets Go! experience. In *Proceedings of INTERSPEECH* (2006), 65–68.
16. Somers, H. Language engineering and the pathway to healthcare: A user-oriented view. In *NAACL Workshop on Medical Speech Translation* (2006), 32–39.
17. Stücker, S., Fügen, C., Kraft, F., and Wölfel, M. The ISL 2007 english speech transcription system for european parliament speeches. In *Proceedings of INTERSPEECH* (2007).
18. Stüker, S., Zong, C., Reichert, J., Cao, W., Kolss, M., Xie, G., Peterson, K., Ding, P., Arranz, V., Yu, J., and Waibel, A. Speech-to-speech translation services for the olympic games 2008. In *Proceedings of MLMI* (2006), 297–308.
19. Turunen, M., Hakulinen, J., Salonen, E.-P., Kainulainen, A., and Helin, L. Spoken and multimodal bus timetable systems: Design , development and evaluation. In *Proceedings of SPECOM* (2005), 389–392.
20. Wahlster, W. *Verbmobil: foundations of speech-to-speech translation*. Springer, 2000.
21. Walker, M. A., Rudnicky, A., Prasad, R., Aberdeen, J., Bratt, E. O., Garofolo, J., Hastie, H., Le, A., Pellom, B., Potamianos, A., Passoneau, R., Roukos, S., Sanders, G., Seneff, S., and Stallard, D. DARPA communicator: cross-system results for the 2001 evaluation. In *Proceedings of ICSLP* (2002), 269–272.