Conversational User Interfaces

Wolfgang Wahlster, German Research Center for AI, DFKI GmbH

Self-service systems, online help systems, web services, mobile communication devices, remote control systems, and dashboard computers are providing ever more functionality. However, along with greater functionality, the user must also come to terms with the greater complexity and a steeper learning curve. This complexity is compounded by the sheer proliferation of different systems lacking a standard user interface.

Conversational user interfaces allow various natural communication modes like speech, gestures and facial expressions for input as well as output and exploit the context in which an input is used to compute its meaning. The growing emphasis on conversational user interfaces is fundamentally inspired by the aim to support natural, flexible, efficient and powerfully expressive means of human-computer communication that are easy to learn and use. Advances in human language technology and intelligent user interfaces offer the promise of pervasive access to online information and web services. The development of conversational user interfaces allows the average person to interact with computers anytime and anywhere without special skills or training, using such common devices as a mobile phone.

Advanced conversational user interfaces include the situated understanding of possibly imprecise, ambiguous or incomplete multimodal input and the generation of coordinated, cohesive, and coherent multimodal presentations. In conversational user interfaces the dialogue management is based on representing, reasoning, and exploiting models of the user, domain, task, context, and modalities. These systems are capable of real-time dialogue processing, including flexible multimodal turn-taking, backchanneling, and metacommunicative interaction.

One important aspect of conversations is that the successive utterances of which it consists are often interconnected by cross references of various sorts. For instance, one utterance will use a pronoun to refer to something mentioned in the previous utterance. Computational models of discourse must be able to represent, compute and resolve such cross references.

Conversational user interfaces differ in the degree with which the user or the system controls the conversation. In directed or menu-based dialogues the system maintains tight control and the human is highly restricted in his dialogue behavior, whereas in free-form dialogue the human takes complete control and the system is totally passive. In mixed-initiative conversational user interfaces, the dialogue control moves back and forth between the system and the user like in most face-to-face conversations between humans.

Four papers in this special issue deal with conversational user interfaces that use speech as the main mode of interaction.

The paper by *Helbig* and *Schindler* discusses state-of-art component technologies and requirements for the successful deployment of conversational user interfaces in industrial environments such as logistics centers, assembly lines, and car inspection facilities. It shows that the speech recognition rate in such environments is still depending on the correct positioning and adjustment of the microphone and discusses the need for wireless microphones in most industrial applications of spoken dialogue systems.

Block, Caspari and *Schachtl* describe an innovative dialogue engine for the Virtual Call Center Agent (ViCA), that provides access to product documentation. A multi-frame based dialogue engine is introduced that supports natural conversations by allowing over-answering and free-order information input. The paper reports encouraging results from a usability test showing a high task completion rate.

The paper by *te Vrugt* and *Portele* describes a tasked-oriented spoken dialogue system that allows the user to control a wide spectrum of infotainment applications, like a hard-disk recorder, an image browser, a music player, a TV set and an electronic program guide. The paper presents a flexible framework for such a multi-application dialogue system and an application-independent scheme for dialogue processing.

Nöth et al. describe lessons learnt from the implementation of three commercially deployed conversational interfaces. The authors propose five guidelines, which they consider to be crucial, when building and operating telephone-based dialogue systems. One of the guidelines concerns the fact that a spoken dialogue system must react fast to any kind of user input, no matter

how long the user utterance is. The authors found that a delay of two seconds or more after the end of the input is likely to confuse the user, because this often misleads the users to repeat or to reformulate their utterance.

Three papers deal with multimodal conversational user interfaces. These systems merge three user interface paradigms, namely spoken dialogues, graphical interfaces, and gestural interaction, to achieve truly multimodal communication.

Wasinger and *Krueger* describe multimodal conversational interfaces for navigation and spatial assistance systems. They discuss the integration of embedded speech recognizers with components for the processing of various types of gestures and handwriting for mobile navigation systems. A blackboard-architecture is described for the fusion of the speech and gesture analysis results.

The paper by *André* and *Rist* presents new research in the area of embodied conversational characters. It presents the trend from face-to-face communication between a single agent and a single user to multi-party multi-threaded interactions between several human and synthetic interlocutors and observers. In addition, it discusses attempts to integrate embodied conversational agents into the user's natural environment.

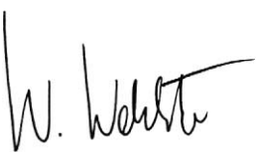
The paper by *Pfleger* and *Alexandersson* discusses the role of non-verbal behavior in multimodal conversational interfaces. In mixed-initiative dialogue systems back-channeling and turn-taking are important means to structure the flow of conversation. It presents a discourse model for conversational dialogue processing emphasizing the importance of interactional information and provides a set of processing rules that enable the treatment of turn-regulating and back-channel behavior together with propositional information.

Currently, we see an increase in the commercial deployment of conversational user interfaces. The first German Voice Award for telephone-based conversational systems was presented in Munich on the 19th of October 2004. As chairman of the Prize Committee, I could award five prizes in the following categories: the best-practice German dialogue system, the most innovative spoken dialogue application, the system with the best dialogue design, the best speech-based enterprise solution, and the best voice-enabled interface to added-value services. 60 deployed and fully operational spoken dialogue systems have been evaluated by the Prize Committee. Two thirds of these systems provide voice-enabled access to self-service systems (e. g. for banking, brokerage, and retail). It is interesting to note that the return of investment (ROI) for these spoken dialogue systems was within 24 months.

I am pleased that the first German Voice Award for the most innovative application has been presented to a system that is described in this special issue: the BERTI system designed and implemented by Sympalog. BERTI is a mixed-initiative dialogue system that deals with the German soccer league. The paper by *Nöth et al.* describes the technology underlying this award-winning system.

Most of the research presented in this volume was initially supported by funding from the German Federal Ministry of Education and Research (BMBF). I would like to thank *Dr. Bernd Reuse*, the Head of the Division for Software Systems at BMBF, for his constant and tireless support of pioneering projects like Verbmobil and SmartKom, that formed the basis for the successful commercial applications described in this special issue.

Finally, I would like to thank the reviewers who provided insightful comments and valuable feedback to the authors of this special issue. Special thanks go to *Dr. Hartmut Raffler* for his initiative to create this special issue.



Wolfgang Wahlster

Prof. Dr. Dr. h.c. mult. Wolfgang Wahlster is the Director and CEO of the German Research Center for Artificial Intelligence (DFKI GmbH) and a Professor of Computer Science at Saarland University, Saarbrücken. He received his diploma and doctoral degree (1981) in Computer Science from the University of Hamburg. He has published more than 160 technical papers and 8 books on intelligent user interfaces. His current research includes multimodal and perceptive user interfaces, user modeling, embodied conversational agents, smart navigation systems, and semantic web services. He is a Fellow of AAAI, ECCAI and GI and a recipient of the Fritz Winter Award (1991), an IST Prize (1995), and the Beckurts prize (2000). In 2001, the President of the Federal Republic of Germany presented the German Future Prize to Prof. Wahlster for his work on language technology and intelligent user interfaces. He was elected Full Member of the German Academy of Sciences and Literature, and Foreign Member of the Royal Swedish Society of Sciences, Stockholm and the German Academy of Natural Scientists, Leopoldina in Halle.
Address: German Research Center for AI, DFKI GmbH, Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany,
E-Mail: wahlster@dfki.de
www.dfki.de/~wahlster/