Directions For Multi-Party Human-Computer Interaction Research

Katrin Kirchhoff
Department of Electrical Engineering
University of Washington, Seattle
katrin@ee.washington.edu

Mari Ostendorf
Department of Electrical Engineering
University of Washington, Seattle
mo@ee.washington.edu

Abstract

Research on dialog systems has so far concentrated on interactions between a single user and a machine. In this paper we identify novel research directions arising from multi-party human computer interaction, i.e. scenarios where several human participants interact with a dialog system.

1 Introduction

Most current work on spoken human-computer interaction (HCI) involves dialog systems. In recent years, spoken dialog systems with system initiative have become more commonplace in commercial telephony applications, and there have been important advances in mixed initiative and multi-modal research systems. Telephone-based systems have made it possible to collect large amounts of human-computer interaction data, which has benefited empirical research as well as methods based on automatic training. In addition, evaluation frameworks have improved beyond the single utterance accuracy measures used a decade ago to dialog-level subjective and quantitative measures (Walker et al., 1998).

As dialog systems have advanced, a new area of research has also been developing in automatic recognition and analysis of multi-party human-human spoken interactions, such as meetings, talk shows, courtroom proceedings, and industrial settings (Cohen et al., 2002). Multi-party interactions pose challenges for speech recognition and speaker tracking because of frequent talker overlap (Shriberg et al., 2001), noise and room reverberation, but they also introduce new challenges for discourse modeling. Until recently, empirical research was only possible using single-speaker and dialog corpora, but now there are many hours of data being collected in multi-talker environments (Morgan et al, 2001; Schultz et al, 2001).

While many challenges remain in dialog systems – from error handling and user modeling to response generation – technology has advanced to the point where one can also envision tackling the combined problem of multi-party human-computer interaction. A key motivation for research in such a domain is supporting human-human collaboration. We envision a scenario where a computer plays a role as a conversational agent, much as in a dialog system, except that it interacts with multiple collaborating humans. The human participants may be at distributed locations, perhaps with small subgroups at each location, possibly with different platforms for input and output. For example, one might imagine a group of people in a facility with high-end computers interacting with workers in the field with lightweight communication clients, using the computer assistant to help gather vital information or help plan a transportation route. A key difference from previous work in such scenarios is the idea of computer initiative. The computer as a participant also significantly changes the focus of research relative to that involved in transcription and analysis of meetings, from work aimed at indexing and summarization to a focus on interaction.

Besides the application-oriented motivation for research on multi-party human-computer interaction, the scenario provides a useful technology pull. In current dialog systems, there is a disincentive to explore user initiative, simply because much better accuracy can be achieved by “controlling” the dialog. However, it would be impractical for a system to try to constrain the inputs from a group of users. Secondly, current dialog systems generally assume a fixed platform, and hence the response generation can be greatly simplified. With varying platforms and participants with different needs, a more complex output rendering strategy will be needed, which will also have implications for future dialog systems as well. In the follow section, we expand on these issues and many more research questions that arise in the context of multi-party HCI research.

2 RESEARCH ISSUES

Some of the most intensively pursued research questions in single-party human-computer interaction are the following: initiative strategies (human vs. system vs. mixed initiative); dialog planning, in particular the possibility of learning probabilistic dialog models from data; handling recognition/understanding errors and other system failures or miscommunications; user modeling, i.e. finding models of interaction patterns specific to certain users in order to adapt the dialog system; and multimodal input/output strategies. A multi-party scenario extends these questions in various interesting ways but also presents entirely new challenges, such as described below.

Initiative. It needs to be asked how human-human communication affects interaction with an automatic dialog system on the one hand, and how the presence of the system influences communication among the human participants on the other. Specifically, how is the frequency and type of each user's interaction with the system determined? Does every user address the system on a "first come, first speak" basis, do users take turns, or do they first communicate among themselves and then interact with the system via a designated "proxy" speaker? How do these different interaction protocols affect the outcome? For instance, communicative goals might be achieved more frequently and more rapidly when the protocol is fixed in advance but users might be more satisfied with a less structured interaction.

Two other questions are closely tied to the issue of interaction and initiative protocols: (a) should the system be an open-mike dialog system, i.e. able to record everything at all times, though responding only to specific events? and (b) are users in the same physical location or are they distributed (e.g. in videoconferencing)? In the case of an open-mike system, special provisions need to be made to distinguish between utterances addressed to the dialog system and utterances that are exclusively part of the human-human interaction. This additional challenge is offset by the possibility of gathering useful background information from the participants' conversation that might enable the system to better respond to queries.

Dialog Modeling. Distributed scenarios, where different subgroups of participants are separated from each other physically, will typically lead to parallel subdialogs evolving in the course of the interaction. In this case the system needs to be able to track several subdialogs simultaneously and to relate them to each other. The possibility of having multiple concurrent subdialogs directly affects the dialog planning component. Different user queries and dialog states might need to be tracked simultaneously, and formal models of this type of interaction need to be established. Recently, probabilistic models of human-computer dialogs have become increasingly popular. The most commonly used paradigm is that of Markov decision processes and partially observable Markov decision processes, where the entire dialog is modelled as a sequence of states and associated actions, each of which has a certain value (or reward) (Singh et al., 1999; Roy et al., 2000). The goal is to to choose that sequence of actions which maximizes the overall reward in response to the user's query. States can be thought of as representing the underlying intentions of the user. These are typically not entirely transparent but only indirectly (or partially) observable through the speech input. Multi-party dialogs might require extensions to this and other modeling frameworks. For instance, it is unclear whether multiple parallel subdialogs can be modelled by a single state sequence (i.e. a single decision process), or whether multiple, partially independent decision process are required. The issue of how to acquire data to train these models is a further problem, since parallel subdialogs tend to occur spontaneously and can often not be elicited in a natural way.

Error handling. The prevention and repair of system errors and miscommunications may take on an entirely new shape in the context of multi-party interactions. First, the notion of what constitutes an error may change since some participants might interpret a particular system response as an error whereas others might not. Second, the inherent susceptibility of the system to recognition and understanding errors will be higher in a group than when interacting with a single user since both the speech input and the interaction style exhibit greater variation. Third, error recovery strategies cannot necessarily be tailored to a single user but need to take into account the input from multiple participants, such as the spontaneous and possibly diverging reactions to a system recognition error. Studies on different reactions to system errors (e.g. building on related studies in single-party HCI (Oviatt et al., 1998)) should be included in a roadmap for multi-party HCI research.

User Modeling. User modeling has recently gained increased importance in the field of dialog systems research, as evidenced by the growing number of dialog-related publications in e.g. the International Conference on User Modeling, and the *User Modeling and Adaptation* journal. When multiple users are present, several concurrent user models need to be established, unless interaction is restricted to a proxy scenario as described above. Here, too, the question is not only what individual models should look like, but also how they can be learned from data. Group interactions are typically dominated by a small number of active speakers, whereas the remaining participants provide fewer contributions, such that multi-party data collections tend to be rather unbalanced with respect to the amount of data per speaker. Furthermore, individual users might behave differently in different scenarios, depending e.g. on other participants in the group.

Flexible "Multi" Input/Output. Research on multimodal input/output for language-based dialog systems is a relatively new field, though many contributions have been made in recent years. Many developments will impact both dialog and multi-party systems, but introducing the multi-party dimension brings further challenges. For example, the problem of coordinating speech and gesture from one person is complicated by increasing the number of people, making the problem of speaker tracking important. For speech input, there are questions of whether to use an open-mike system, as mentioned earlier, but there may also be different requirements for participants with distributed platforms/locations (e.g. noisy environments may require push-to-talk). One could envision haptic devices controlled simultaneously be multiple users. On the output side, there is a problem of incorporating backchannels and visual evidence of attentiveness (equivalent to head nods), as well as turn-taking and interruption cues for coordinating with other human speech. Coordination of different output modalities faces an added challenge when some platforms/environments preclude use of all modalities. Considering language alone, the response generator must provide alternatives depending on whether voice output is available and on display size (i.e. how much text and/or visual aids can be included). User and context modeling should also impact response generation.

3 INFRASTRUCTURE

In order to study the research questions addressed above we need appropriate resources. Currently, no publicly available corpus of multi-party human-machine interactions exists. Several corpora of human-human communication are available and may be used to study phenomena such as negotiation of turn-taking but are clearly not sufficient to support work on multi-party HCI.

Data collection mechanism. It would be desirable to collect several corpora of multi-party human-machine communication, representing different scenarios, e.g. corporate meetings, remote collaboration of scientific teams, or, remaining closer to existing scenarios, collaborative travel planning of business partners. Care should be taken to collect data from groups of various sizes, co-located as well as distributed teams, technologically experienced vs. inexperienced users, different input modalities, and teams working on a variety of different tasks. Ideally, data collection should be coordinated across different sites with complementary expertise. Data collections should be made available publicly, e.g. through LDC. Existing multi-party recording facilities (such as instrumented meeting rooms) could be leveraged for this effort.

New Evaluation Paradigms. One of the most important research questions is how to evaluate the success of multi-party HCI. Can we build on existing frameworks for single-person dialogs? For example, can we extend the Paradise framework (Walker et al., 1998) by introducing new quantitative measures (such as speaker tracking error, computer interruption rate) and designing group questionnaires or averaging responses to individual questionnaires? As in dialog system research, component-level evaluations will continue to be a key driver of research progress, but a multi-party system would likely include new components relative to a dialog system. For example, a natural task for the computer might be information retrieval (IR), in which case there measures from the IR community would be relevant. Additionally, we can incorporate insights from more general (i.e. not necessarily speech-specific) evaluation frameworks for collaborative systems (Drury et al, 1999; Damianos et al, 2000), which take into account factors such as group size, social interaction parameters, and collaborative tasks.

Multi-party HCI represents a substantial step beyond current research, but it is an important challenge that will drive new ideas in many areas. Since multi-party HCI is fundamentally about collaboration, it is an ideal problem for fostering the type of multi-site and multi-discipline interactions that will advance human communication

References

P.R. Cohen, R. Coulston, and K. Krout. 2002. Multiparty multimodal interaction: A preliminary analysis. In *Proc. of ICSLP*, pages 201–204.

L. Damianos et al. 2000. Evaluating multi-party multi-modal systems. Technical paper, The MITRE Corporation.

J. Drury et al. 1999. Methodology for evaluation of collaborative systems, v.4.0. http://www.nist.gov/nist-icv.

N. Morgan et al. 2001. The meeting project at ICSI. In *Proc. of HLT*, pages 246–252.

S. Oviatt, G.A. Levow, E. Moreton, and M. MacEachern. 1998. Modeling global and focal hyperarticulation during human-computer error resolution. *JASA*, 104(5).

N. Roy, J. Pineau, and S. Thrun. 2000. Spoken dialogue management using probabilistic reasoning. In *Proc. of ACL*.

T. Schultz et al. 2001. The ISL meeting room system. In *Proc. Workshop on Hands-Free Speech Communication (HSC-2001)*, Kyoto Japan.

E. Shriberg, A. Stolcke, and D. Baron. 2001. Observations on overlap: Findings and implications for automatic processing of multi-party conversation. In *Proc. EUROSPEECH*, pages 1359–1362.

S. Singh, M. Kearns, D. Litman, and M. Walker. 1999. Reinforcement learning for spoken dialog systems. In *Advances in Neural Processing Systems*, volume 12.

M. Walker et al. 1998. Evaluating spoken dialogue agents with PARADISE: Two case studies. *Computer Speech and Language*, 12(3):317–348.