

Natural Language in Computer Human-Interaction

A CHI 99 Special Interest Group

Nancy Green and David Novick

Fifty-three people from across the world participated in the CHI 99 special interest group on Natural Language in Computer Human-Interaction. The SIG's main goal was to provide an opportunity for CHI99 attendees from two research communities, natural language processing (NLP) and human-computer interaction (CHI), to discuss issues of mutual interest. The SIG embraced natural-language interfaces of all kinds, including text, spoken and multi-modal interaction. This report includes the results of e-mail discussions following up on the SIG itself.

The NL SIG's motivation was that the growing interest in human-computer interfaces that use spoken or written natural language in some way is leading researchers and practitioners who work on these interfaces to find that the two fields of research, CHI and NLP, are complementary and converging. In the CHI research community, there have been investigations on a number of related issues such as usability of text and graphics in on-line documentation, (e.g., Landauer et al.; Brockmann, 1986), hypertext, (e.g., Chen & Rada, 1996), spoken-dialogue interfaces, and (e.g., Yankelevich, Levow, & Marx, 1995; Hansen, Novick & Sutton, 1996; Walker et al., 1998), and language/audio resources. (e.g., Arons, 1993). In the NLP research community, there is increasing interest in use of natural language in intelligent multimodal and multimedia interfaces, e.g., the International Symposium on Spoken Dialogue (ISSD-96), the COOP 98 Workshop on the Use of Herbert H. Clark's Models of Language Use for the Design of Cooperative Systems, and the 1998 AAAI Workshop on Representations for Multi-Modal Human-Computer Interaction.

Topics

The participants proposed a number specific discussion topics related to the SIG's main theme. Issues of interest to participants included:

The CHI-NL community. Why is there so little speech and natural language at CHI, and how can we bring more natural language into CHI? Conversely, what could the Association for Computational Linguistics do to be more CHI-oriented?

State of the art. How realistic is it to have a spoken-language system for a consumer-level television set, particularly as an interface to the Web? What is the state of the art for speech recognition and spoken-language systems? What will the state of the art be in five years?

Technical issues. How can we make spoken-language systems more non-intrusive and usable? What are the roles of verbal, auditory and visual feedback? Can we differentiate dialogue styles? How can we support interaction in multiple languages and regional dialects?

Discussion

The participants rated, combined and selected three particular issues for discussion.

The SIG first discussed the combined topic of (a) whether it is realistic to have a consumer television with a spoken-language interface to the Web, and (b) the state of the art. Participants indicated that speaker-independent speech recognition (SR) is feasible now. For example, telephone companies have SR systems for long-distance billing. But in

the application of command and control of a television in the home, the acoustic environment is important. For a Web interface, barge-in needs to be supported. Participants noted that a number of speech-enabled Web browsers already exist. One factor for the effectiveness of such systems is whether the size of vocabulary is limited. There are off-the-shelf products available for developing command-and-control applications, including the OGI toolkit.

More fundamentally, a key problem is that many spoken-language systems have no real understanding of unrestricted dialogue above and beyond speech-level. This affects usability. So in considering what kinds of task for which current SR systems are useful, one must appreciate the difference between voice-control of buttons and true dialogue. Current spoken-language systems developed in North America include Nuance/Altech, the systems produced by Victor Zue's group at MIT (which can answer questions like "What is the weather in Boston"), the DARPA Communicator (airline reservations), W3C (voice browser standardization), and C-star (NL speech translation). Systems developed in Europe include LIMSI's RailTel system, the SUNDIAL system, the Danish Dialogue System, and the Verbmobil (translation for business phone calls).

The second major topic discussed by the participants was the usability of natural-language interfaces. One element that participants mentioned as being usable now was verbal feedback. More generally, usability is affected by the influence of task on the problem. For example, a train reservation system needed to understand certain kinds of referring expressions that were not needed, in contrast, in an airline reservation domain. Similarly, Nils Dahlback found that a local bus traffic information system needs a more sophisticated system for managing referring expressions than does a train reservation system. Another factor affecting usability is the developer's choice of modality. Is language being used for input, output, or both? Does the system use speech or text? One sticking point for usability of SR systems is that user control of speech-recognition hardware is difficult. Finally, the participants discussed the usability implications of interfaces using multiple languages for example, to support a 40-language interface, rather than using 40 times the storage, instead modularize and download only what needed for the language in use.

The third major topic discussed by the participants was the relation of the CHI, natural-language processing (NLP), and spoken-language systems (SLS) communities. In considering the relationships among these communities, one should bear in mind that all of these are tied to dialogue. That is, the practice and research issues involve not only speech but speech-based dialogue. Candy Sidner, among others, pointed out that in many cases there is a lack of understanding of the difference between having a speech input to something (which is basically using isolated commands) and interacting in a dialogue. Indeed, one of the reasons for the lack of interest in spoken-dialogue systems in the CHI community may be a lack of awareness of the difference between

the two cases. We discussed whether CHI had an outright hostility to papers on interfaces using text or speech. Some natural-language and dialogue papers have been rejected on the grounds that they weren't "human-computer interaction."

Questions came from the CHI "side" as to where CHI members can find out what is state-of-the-art in natural-language processing and spoken-language systems, not necessarily in terms of research frontiers but more in terms of what applications are feasible. It turned out that there was a corresponding gulf between the natural-language and speech-recognition communities. Most speech people don't attend the annual conference of the Association for Computational Linguistics (ACL), and most NLP people don't attend speech conferences. And members of neither community tend to attend CHI, partly because of repeated rejections from CHI ("not a CHI paper"). As an overall matter, we of the natural-language interaction community represented at the SIG have a problem in that our research is coherent but we have difficulty getting papers accepted. We have to submit them across their component disciplines (CHI, NLP, SLS) because there is not yet a single audience for the work as a whole.

The participants expressed interest in having CHI tutorials and/or workshops on natural-language and spoken-language interfaces and systems. We discussed the relationship between CHI and SIGDIAL, the ACL's special interest group on dialogue and discourse. Perhaps common activities could be encouraged. Certainly, members of SIGCHI and of ACL need to raise awareness of each other. Their common interests in natural language include usability and cognitive/interactional models of language communication.

Action Items

The SIG concluded with suggestions of action items to address the issue of the fragmented NL-CHI community. One suggestion was that members of the community organize tutorials and/or workshops at CHI2000, and a workshop on NL-CHI was prepared and submitted in September.

Participants suggested that one way to increase the number of NL papers accepted at future CHI conferences would be by volunteering for reviewing, by participating in the program committee, and by chairing paper sessions. And to encourage increased submissions of NL-CHI papers, the NL-CHI community needs to solicit papers, in part by promoting SIGCHI and the CHI conferences in the natural-language and spoken-language systems communities. On a practical basis, the SIG organizers plan to continue to support and update the SIG's Web site at <http://www.cs.utep.edu/novick/chi/nlsig.html>. Updates include a list of SIG participants and other interested professionals, links to the ACL/SIGDIAL Web pages, and links to tools and reviews of tools. In the future, the community may want to create other means of communication, possibly through a CHI-sponsored newsgroup.

References

Arons, B. (1993). Speech Skimmer: Interactively skimming recorded speech. Proceedings of UIST 93, 187-196.

Brockman, John R. Writing Better Computer User Documentation: from Paper to Online. John Wiley and Sons, Inc., New York, 1986.

Chen, C., and Rada, R. Interacting with Hypertext: A Meta-Analysis of Experimental Studies. Human-Computer Interaction, 1996, v. 11, 125-156.

Hansen, B., Novick, D., and Sutton, S. (1996). Systematic design of spoken prompts, Conference on Human Factors in Computing Systems (CHI'96), Vancouver, BC, April, 1996, 157-164.

Landauer, T., Egan, D., Remde, J., Lesk, M., Lochbaum, C., and Ketchum, D. Enhancing the Usability of Text through Computer Delivery and Formative Evaluation: the Super-Book Project, in McKnight, C., Dillon, A., and Richardson,

J. (eds), Hypertext: A Psychological Perspective, New York, Ellis Horwood, 1993, 71-136.

Walker, M., Fromer, J., Di Frabrizio, G., Mestel, C., and Hindle, D. (1998). What can I say?: Evaluating a spoken language interface to email, Human Factors in Computing Systems (CHI 98), Los Angeles, CA, April, 1998, 582-589.

Yankelovich, N., Levow, G., and Marx, M. Designing SpeechActs: Issues in speech user interfaces, Proceedings of Computer Human Interaction, 1995, 369-376.

Author

Nancy Green, University of North Carolina at Greensboro, nlgreen@uncg.edu

David G. Novick, University of Texas at El Paso, novick@cs.utep.edu

