

1 Research Interests

My research interests are in dialogue systems and computer-assisted learning. Within dialogue systems I focus on improving the quality of the human-computer interaction at the level of dialogue behaviors. This includes areas such as increasing the interactivity of dialogue systems with new methods for **turn-taking, incremental understanding, interleaved speech and actions**, as well as the integration of **visual and dialogue context** in processing. Within computer-assisted learning I focus on **language learning** (reading and writing) and **task rehearsal and performance**.

1.1 Past, Current and Future Work

I have a particular interest in interactivity, such as turn-taking and incremental understanding.

I invented a general architecture to support full-duplex human-computer conversation, where either, neither, or both the human and the computer could speak at the same time. (Vs. only one at a time.) This architecture enabled the introduction of novel turn-taking behavior (such as backchanneling and interruption) in Project LISTEN's Reading Tutor (Mostow et al., 2003), which listens to children read aloud and helps them learn to read.

I then devised methods so a dialog system could respond to commands given while the computer was still speaking. For example, during the development of NASA's Clarissa astronaut assistant (Aist et al., 2004): if the user broke in and said "Speak up", then Clarissa could immediately increase the output volume in response.

I then began working on parsing techniques, system architectures, and semantic representations to allow the system to begin actions while the user's utterance is still unfolding (Aist et al., 2006). For example, given a square decorated with a diamond, a region named Central Park, and a flag as a landmark, assume the user said "First move a large square with a diamond on the corner into Central Park right on top of the flag." By the word "right" we know the large square goes into Central Park, even if we don't know exactly where, and so we can begin the movement - adjusting the exact target later.

I have also developed and co-developed several novel techniques of experimental methods in interactive sys-

tems: (a) embedded experiments for fine-grained evaluation of student learning and system performance; (b) mining human-human dialogs for alternate phrasings in order to enrich human-computer dialog; and (c) system-user-expert dialogs for expanding the expertise of an existing system (Aist, 2004).

2 Applications and Challenges in Spoken Dialog Systems Research

Many possible applications present new and exciting opportunities and challenges for dialogue systems, such as computer gaming, personal assistant(s), and multiperson dialogues. Here, however, I present an application area unusual in that it is process-oriented rather than an end in itself: the use of spoken dialogue systems to enable scientific research in other fields.

Using dialogue systems as an enabler of research could take several forms, such as (a) dialogue systems that conduct surveys or interviews to gather data from study participants, (b) dialogue systems that are constructed as working models of human cognition in a domain, or (c) dialogue systems that are constructed in order to interact with users to resolve scientific questions in the domain of conversation.

What sort of impact could dialogue systems have on the research process? Dialogue systems that conduct surveys could gather free-response data that is qualitatively richer than that collected by "press 1 for yes" touch-tone surveys. Dialogue systems could also reach people in non-telephone settings such as mall kiosks. And, in a laboratory setting, it is possible that people might feel more comfortable when answering uncomfortable questions if they are talking to a computer rather than another person.

Dialogue systems as models of human cognition would allow human factors analysis of tasks that involve conversation interleaved with action. A key example is air traffic control; other examples include search-and-rescue, driving while following directions, and surgery. As a predictive model a dialogue system modeling an air traffic controller or a surgeon would provide hypotheses about how a real person would perceive and respond in various situations - both the ordinary and the unusual, including problems that arise during the task.

Dialogue systems that are constructed so that they engage in conversation with users in order to answer scientific questions in the domain of conversation are currently most prevalent in the field of tutorial dialogue. (In some sense any tutorial dialogue system inherently poses the question “what sort of human learning takes place during the interactions this dialogue system provides?”.) Other possible scenarios include listening in to human-human conversation, and then interjecting clarification questions or other probes in order to get a better picture of what is happening in the human-human dialogue.

What questions need to be answered in order to realize spoken dialogue systems that help enable the research process? For spoken dialogue systems that take surveys, one key aspect is data security - for example, having a model of who should have access to what other data about whom. It would be potentially a powerful tool for social research to have a dialogue system that could ask people about their opinions of other people’s behavior - for example, asking each study participant to compare their behavior to a similar participant who made different decisions - but such questions and their answers would need to take into account concerns such as privacy.

For dialogue systems as working cognitive models, we would need detailed accounts of how language and action relate in the domain of study. (At least to where the model-building exercise was based on rich enough abstract constructs to yield initial testable hypotheses.)

Finally, dialogue systems that are constructed to help answer scientific questions in the domain of conversation need to be able to keep careful records of the conversation itself as well as other non-conversational factors that might weigh on the analysis - user demographics, time of day, presence of other people (in a field study), and so forth. In addition, such systems will need to be able to conduct randomized controlled trials - in some cases varying behavior(s) between subjects, and in other cases varying behavior(s) within one subject’s experience according to the independent variables under investigation.

3 Suggestions for Discussion

Here are three suggestions for discussion. (The first two elaborate on questions posed in the Call for Papers.)

Use and uselessness of user simulations: Best practices for building and evaluation user simulations?

User simulations offer a way to perform exploratory research much more quickly than live user studies. Yet how well does the simulation match actual user behavior?

Dialogue and question answering: How can dialogue research help with interactive question answering?

Many commercial search engines are close to stateless; how can the use of dialogue history help?

Input/Output Asymmetry

In some scenarios, speaking to a computer may be very natural and useful, while listening to the answer may not be. For example, spoken access to a database of maps might involve spoken input and text-plus-graphics output. (You could say “How do I get to the nearest ATM?” and be presented with a map of the directions.) In other scenarios, speaking to a computer might not be natural or useful, while listening to answers might be preferred. For example, accessing music on a satellite radio might involve button-oriented input and spoken output. (You could use the up/down buttons to select a new station and then hear the new station name in a voice-over as the sound fades in: “Easy Listening Radio 41 - ”). Where do such scenarios fit into the research and development of spoken (and multimodal) dialogue systems?

References

- G.S. Aist, S. Stoness, and J. Allen. 2006. Steps toward incremental semantics for spoken dialog systems. *The Third Midwest Computational Linguistics Colloquium*.
- G.S. Aist. 2004. Three-way system-user-expert interactions help you expand the capabilities of an existing spoken dialogue system. *International Conference on Spoken Language Processing*.
- G.S. Aist, D. Bohus, B. Boven, E. Campana, S. Early, and S. Phan. 2004. Initial development of a voice-activated astronaut assistant for procedural tasks: From need to concept to prototype. *Journal of Interactive Instruction Development*, 16(3): 32–36.
- J. Mostow, G. Aist, P. Burkhead, A. Corbett, A. Cuneo, S. Eitelman, C. Huang, B. Junker, M.B. Sklar, B. Tobin. 2003. Evaluation of an automated Reading Tutor that listens: Comparison to human tutoring and classroom instruction. *Journal of Educational Computing Research*, 29(1): 61–117.

Biographical Sketch

Gregory S. Aist works in dialogue systems and computer-assisted learning. At Carnegie Mellon he worked on Project LISTEN’s Reading Tutor, which listens to children read aloud and helps them learn to read. The Reading Tutor was the first intelligent tutoring system to perform anywhere near the level of expert human tutors, in this case performing comparably in terms of vocabulary learning for third graders (the subject of his dissertation.) At RIACS/NASA Ames, he founded the Clarissa Astronaut Assistant Project which talks astronauts through procedures; Clarissa became the first dialogue system used in outer space. He has also held positions at Macquarie University, the MIT Media Lab, and most recently the University of Rochester with the TRIPS group. He is also the founder of the *Journal of Dialogue Systems*, the first journal in the field of dialogue systems.