Adaptive Machinery to Support Natural Conversations

Tim Paek

Department of Psychology
Stanford University
Stanford, CA 94305
paek@psych.stanford.edu

Eric Horvitz

Adaptive Systems and Interaction
Microsoft Research
Redmond, WA 98052
horvitz@microsoft.com

Extended Abstract

Real-time adaptation and ongoing learning promise to provide computational machinery to support fluid conversational interaction. We will focus on representations and inferential strategies that can provide real-time adaptation to multiple characteristics of a conversation. The work approaches conversation from a perspective of inference and decision making under uncertainty. Automated methods for engaging users in conversations must address critical uncertainties about users' attentional status, utterances, and goals.

Although uncertainties arising in conversation with a computer may pertain largely to the special context of interacting with a computer, they are not uncommon in conversations among people. Limitations of speech recognition hardware and software may severely limit the abilities of computers to interpret a user's utterances with certainty. Furthermore, automated inference procedures and representations may not be able disambiguate a user's intentions and needs. However, grappling with uncertainties about someone's utterances and about their associated goals is also common in human to human conversations. People frequently employ a wide range of strategies to resolve uncertainty and repair possible misunderstandings as part of a collaborative process known as *grounding*—or converging on a shared understanding. Although the precise nature and distribution of the uncertainties may shift in moving from human--human to human--computer conversation, we believe that there is great opportunity to develop representations and inference strategies that allow computers to engage users in a natural, expected manner in the course of converging on a shared understanding during conversations.

In the *Conversational Architectures* project, we have been exploring methods for performing inference under uncertainty about the best actions to take to resolve key uncertainties and ambiguities in conversational systems. Our research centers on the development of principles and architectures for engaging in grounding in conversation. Our approach is motivated by the intuition that modeling the way people in conversation employ natural, compensatory strategies for resolving ambiguity and pursuing a shared understanding can provide adaptive machinery for grappling with uncertainties associated with the limited abilities of computers, in a manner akin to the familiar experience of conversing with a person of poor language or hearing skills.

In the extended abstract, we will briefly introduce four levels of analysis and describe representations and inference strategies for managing uncertainty within and between levels (Paek & Horvitz, 1999). The operation of the methods will be highlighted by reviewing interactions drawn from conversation between a user and a prototype system named the *Bayesian Receptionist* (Horvitz & Paek, 1999). The *Bayesian Receptionist* harnesses Bayesian inference, natural language parsing, and speech recognition (Heidorn, 1999) to control dialog about tasks usually handled by receptionists at the front desks of buildings at the Microsoft corporate campus.

Our framework is built upon earlier research that has investigated how people collaboratively contribute to a conversation at successive levels of mutual understanding through grounding (Clark & Schaefer, 1987, 1989). While researchers have examined the relationship of these multiple levels with miscommunication (Brennan & Hulteen, 1995; Dillenbourg et al., 1996; Traum, 1994; Traum & Dillenbourg, 1996, 1998), relatively little work has focused on exploiting uncertainty; for example, by explicitly quantifying uncertainty in terms of probabilities at each level. The framework we present broadens the scope of previous models of grounding and referential expressions (Edmonds, 1993; Heeman, 1991; Heeman & Hirst, 1992; Hirst et al., 1994) by highlighting the efficacy of Bayesian networks and decision theory to reason about uncertainty before and during misunderstanding. Furthermore, the introduction of decision theory allows systems to use expected utility to provide fine-grain, context-sensitive guidance of compensatory measures, rather than relying solely on ad hoc procedures (Brennan, 1998).

Four Levels of Representation and Analysis

Previous attempts to model dialog as a joint activity have focused primarily on the coordination of communication based on propositional beliefs (Cohen & Levesque, 1991, 1994; see Haddadi, 1995 for a review). The logic-based approach to joint activity overlooks critical aspects of joint coordination in dialog that span several different levels of mutual understanding. For example, speakers often repeat themselves if they believe they were not heard since, in a joint activity, it is not enough to just produce utterances; speakers must check that their utterances were attended to and that listeners are still engaged in the activity at hand. Taking inspiration from Clark (1996), we consider four levels of grounding in the pursuit of mutual understanding, displayed in Figure 1.

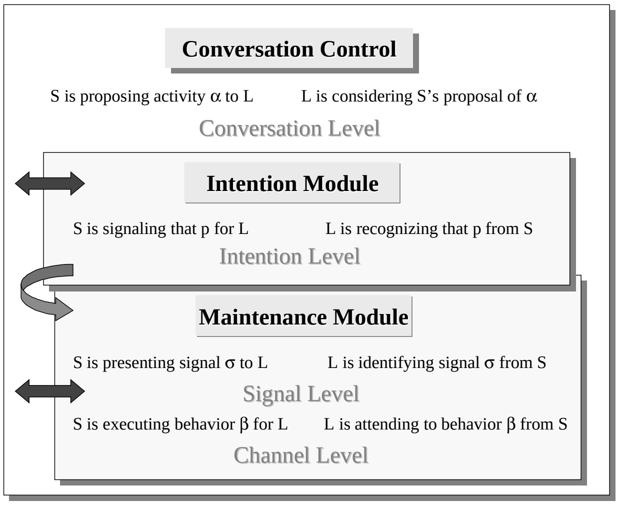

Figure 1. Four levels of representation for inference and decision making under uncertainty in conversation.

At the most basic level, which we denote as the *channel level*, a speaker *S* attempts to open a channel of communication by executing behavior β, such as an utterance or action, for listener *L*. However, *S* cannot get *L* to perceive β without coordination: *L* must be attending to and perceiving β *precisely* as *S* is executing it.

At the next higher level, the *signal level*, *S* presents β as a signal σ to *L*. Not all behaviors are meant to be signals, as for instance, the behavior of a listener scratching an itch during a conversation is irrelevant to the content of the interaction. Hence, *S* and *L* must coordinate on what *S* presents with what *L* identifies.

The *intention level* refers to the task of understanding the semantic content of signals. To date, research on conversational systems has been focused almost entirely on the intention level. At this level, *S* signals some proposition *p* for *L*. What *L* recognizes to be the goal of *S* in signaling σ is *how L* will arrive at *p*. Note that the signal σ is different from the *goal* of *S* in using σ (*e.g.*, in indirect speech acts). By focusing on the goals of *S*, the intention level treats the "speaker's meaning" (Grice, 1957) as primary. *S* cannot convey *p* through σ without *L recognizing* that *S* intends to use σ. This again takes coordination.

Finally, at the *conversation level*, *S* proposes some joint activity α which *L* considers and takes up. A proposal solicits an expected response defined by α. For example, in an indirect speech act such as "I have to go to the North Campus," meaning "Please call a shuttle to the North Campus," *S* is proposing an activity for *S* and *L* to carry out jointly—namely, that *S* gets *L* to call a shuttle. *S* cannot get *L* to engage in the activity without the coordinated participation of *L* in calling a shuttle.

In short, all four levels require coordination and collaboration in order to achieve mutual understanding.

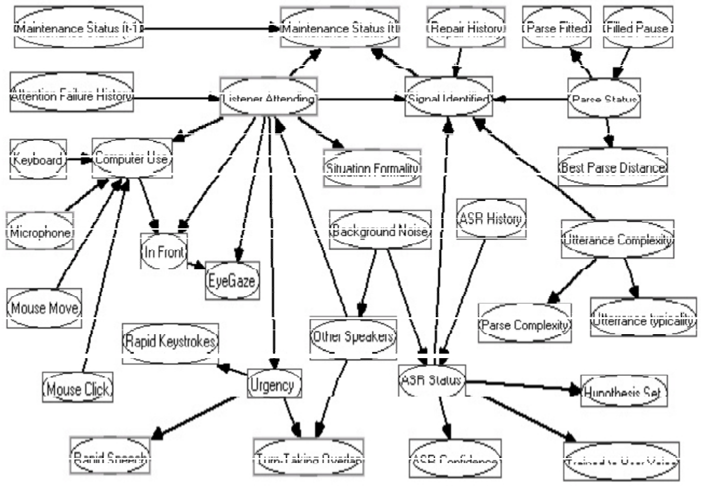

Figure 2. A slice of a larger temporal Bayesian network for reasoning about misunderstanding in the Maintenance Module, representing variables considered at a particular time period.

Dialog Action as Inference and Action Under Uncertainty

Unlike previous models of grounding that represent multiple levels, a decision-making framework allows for uncertainty about what level of misunderstanding a system may be encountering. Rather than having a problem at just one level, and taking action for that level only, a system may be uncertain about which level to investigate, as well as what the costs and benefits of exploring different actions at different levels may be. This problem is especially serious for modular dialog systems that integrate information from a wide variety of component technologies.

We employ Bayesian reasoning and expected value decision making to identify ideal actions in dialog, taking into consideration uncertainties about communication fidelity and meaning, and the potentially varying costs and benefits of alternate actions taken under these uncertainties. We compute the likelihood of states of interest that we cannot observe directly with Bayesian networks. Bayesian networks have been used previously in several user modeling projects (*e.g.,* see Conati et al., 1997; Horvitz, 1997; Horvitz et al., 1998).

As shown in Figure 1, our approach can be viewed as two modules within a larger control subsystem. The *Maintenance Module* handles uncertainty about signal identification and channel fidelity. A Bayesian network for the Maintenance module is shown in Figure 2. The Maintenance Module supports the *Intention Module*, which handles uncertainty about the recognition of user goals from signals. Surrounding both Modules is the *Conversation Control* subsystem which handles uncertainty about the status of the joint activity, Gricean maxims (Grice, 1975), common ground (a shared knowledge base for dialog), and other higher-level dialog events relevant to the joint activity. As represented by the arrows, the Conversation Control subsystem continually exchanges information with both modules and decides where to focus on grounding mutual understanding. The Conversation Control subsystem also adjusts costs or utilities based on records it keeps of conversation level observations, such as the number of questions asked and the number and recency of repair sequences engaged about speech recognition.

For each component of the infrastructure, we exploit the power of value of information (VOI) analysis to identify the best evidence to observe in light of inferred probabilities. To compute VOI, the system calculates for every observation, the expected utility of the best decision associated with each value the observation may take on. The analysis sums the expected utility for each value, weighted by the probabilities of observing different values should an observation be made (see Horvitz, Breese, & Henrion, 1988 for background and details on computation of VOI). Once it recommends which a piece of evidence to observe, a query frame for soliciting that information is used. For example, a user may approach the *Bayesian Receptionist* and ask, "I uh … I need … how do I get to building 25?" The system computes a probability distribution over the goals of the user and determines that two goals, SHUTTLE (i.e., a request for a shuttle) and DIRECTIONS (i.e., asking for directions), are very close in likelihood. Since the maximum likelihood is less than a threshold for checking the maintenance level, the system performs inference over a Bayesian network in the Maintenance Module. The results are displayed in the Figure 3. Here, the most likely state of the maintenance level is CHANNEL AND NO SIGNAL, an apt assessment given a natural language parser that is not equipped to handle restarts, as in the user utterance. This information is passed via the Conversation Control subsystem to the Intention Module which now evaluates the costs and benefits of selecting various types of repair measures.

Using VOI to consider the best observations to make, the system recommends asking a question that tries to observe the word "directions," or any related terms, to discriminate between between the goals of SHUTTLE and DIRECTIONS. Since the most likely maintenance state is CHANNEL AND NO SIGNAL, the system chooses a query frame that specifies a possible misunderstanding at the maintenance level. The output is the combination, "I'm sorry, I may not have heard you properly. Did you want directions?"

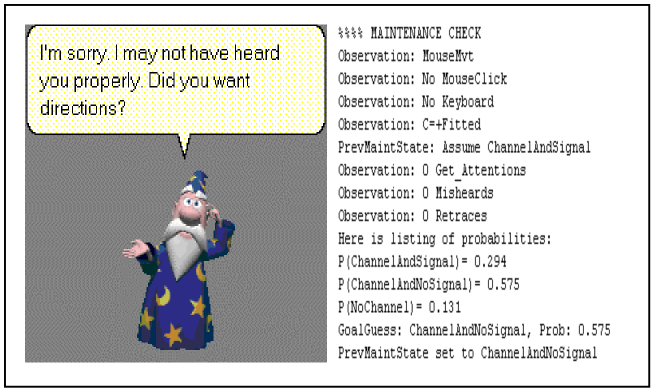

Figure 3. After performing a value-of-information analysis, the *Bayesian Receptionist* decides to ask about directions.

Making Decisions about Repairs in Conversations

The *Conversation Control* subsystem manages uncertainty at the conversation level while continually exchanging information with both the Maintenance and Intention Modules. Taking into account dialog and user event history, the Conversation Control subsystem evaluates when a possible misunderstanding has occurred and how to collaboratively resolve that with users. Conversation Control relies on a sequence of decision analyses to identify the best dialog actions and gestures to interleave with user inputs, responses, and behavior. We decompose the decision making into two problems using influence diagrams (Howard & Matheson, 1984). One influence diagram is used to make decisions about when to repair, and the other is invoked to select the best repair strategy.

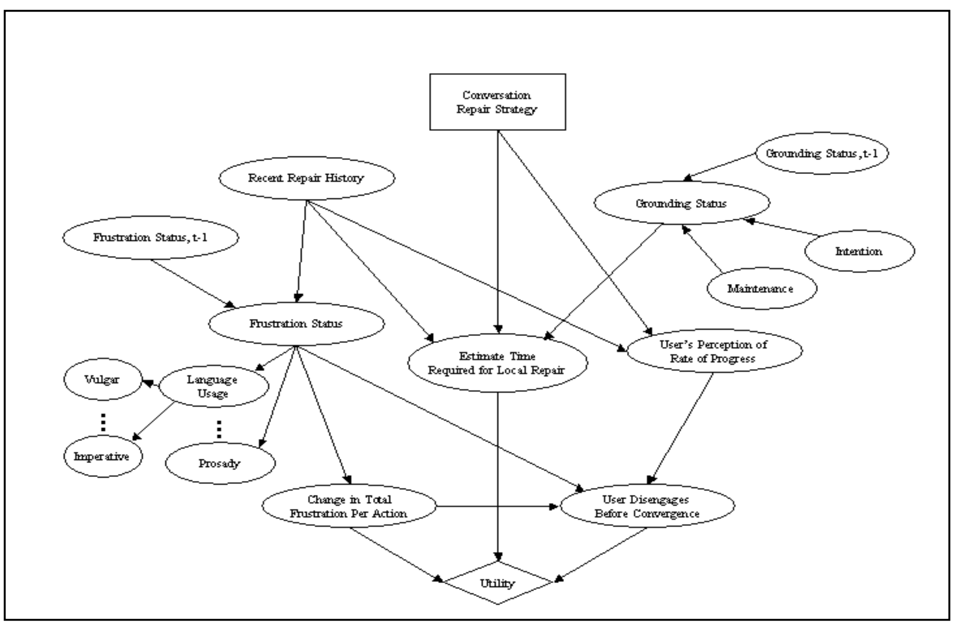

Figure 4. An influence diagram in the Conversation Control subsystem for making decisions about the best conversation repair strategy to select given uncertainty at the different levels of grounding.

Figure 4 shows an influence diagram for deciding which conversational repair strategy to adopt. The two nodes INTENTION and MAINTENANCE represent uncertainty from their respective modules and is evaluated in conjunction with prior history to compute a probability distribution over the four levels of GROUNDING STATUS and an extra state indicating that the dialog is moving along smoothly without high uncertainty in any of the four levels.

The influence diagram considers three primary factors contributing to the utility of any particular repair strategy: estimated time to complete the strategy (ESTIMATE TIME REQUIRED FOR LOCAL REPAIR), a possible change in the frustration status of the user (CHANGE IN TOTAL FRUSTRATION PER ACTION), and the danger of having the user quit the activity altogether (USER DISENGAGES BEFORE CONVERGENCE). The last variable was treated explicitly to underscore the high cost of pushing the user too far. Repair strategy utilities are adjusted by RECENT REPAIR HISTORY, which considers the success rates of the strategies. As indicated in the decision model, the variables FRUSTRATION STATUS and GROUNDING STATUS are influenced by the status of user frustration and grounding at earlier time periods.

In the full AAAI Spring Symposium paper, we will delineate the actions and strategies included in the influence diagrams of the Conversation Control subsystem, and describe in detail the functioning of the multilevel framework for grounding.

References

Brennan, S.A. 1998. The grounding problem in conversations with and through computers. In *Social and Cognitive Approaches to Interpersonal Communication,* 201-225. Lawrence Earlbaum Associates

Brennan, S.A. & Hulteen, E. 1995. Interaction and feedback in a spoken language system: A theoretical framework. *Knowledge-Based Systems* 8: 143-151.

Clark, H.H. 1996. *Using Language*. Cambridge University Press.

Clark, H.H. & Schaefer, E.F. 1987. Collaborating on contributions to conversations. *Language and Cognitive Processes* 2(1): 19-41.

Clark, H.H. & Schaefer, E.F. 1989. Contributing to discourse. *Cognitive Science* 13: 259-294.

Cohen, P.R. & Levesque, H.J. 1991. Teamwork. *Nous*, 25(4): 487-512.

Cohen, P.R. and Levesque, H.J. 1994. Preliminaries to a collaborative model of dialogue. *Speech Communication* 15: 265-274.

Conati, C., Gertner, A., Van Lehn, K., and Druzdzel, M., 1997. Online student modeling for coached problem solving using Bayesian networks*. Proc. of the Sixth International Conference on User Modeling*, 231-242. Springer-Verlag.

Dillenbourg, P., Traum, D. & Schneider, D. 1996. Grounding in multi-modal task-oriented collaboration. *Proc. of the EuroAI & Education Conference*.

Edmonds, P.G. 1993. A computational model of collaboration on reference in direction-giving dialogues, M.Sc. thesis, Dept. of Computer Science, University of Toronto.

Grice, H.P. 1957. Meaning. *Philosophical Review* 66: 377-388.

Grice, H.P. 1975. Logic and conversation. In *Syntax and Semantics 3: Speech Acts*, 41-58. Seminar Press.

Haddadi, A. 1995. *Communication and Cooperation in Agent Systems: A Pragmatic Theory*. Springer-Verlag.

Heeman, P. 1991. A computational model of collaboration on referring expressions. M.Sc. thesis, Dept. of Computer Science, University of Toronto.

Heeman, P., and Hirst, G. 1992. Collaborating on referring expressions (TR 435). Department of Computer Science, University of Rochester, Rochester NY.

Heidorn, G. 1999. Intelligent writing assistance. In *A Handbook of Natural Language Processing Techniques*. Marcel Dekker.

Hirst, G., McRoy, S., Heeman, P., Edmonds, P., & Horton, D. 1994. Repairing conversational misunderstandings and non-understandings. *Speech Communication* 15: 213-229.

Horvitz, E. 1997. Agents with beliefs: Reflections on Bayesian methods for user modeling. *Proc. of the Sixth International Conference on User Modeling*, 441-442. Springer-Verlag.

Horvitz, E. & Paek, T. 1999. A computational architecture for conversation. *Proc. of the Seventh International Conference on User Modeling*, 201-210. Springer Wien.

Horvitz, E., Breese, J., and Henrion, M., 1988. Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning*, 2: 247-30.

Horvitz, E., Breese, J., Heckerman, D., Hovel, D., Rommelse, D., 1998. The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. *Fourteenth Conference on Uncertainty in Artificial Intelligence,* 256-265. Morgan Kaufmann.

Howard, R.A. & Matheson, J.E. 1984. Influence diagrams. In R.A. Howard & J.E. Matheson (Eds.), *The Principles and Applications of Decision Analysis*, Menlo Park, CA: Strategic Decisions Group.

Paek, T. & Horvitz, E. 1999. Uncertainty, utility, and misunderstanding: A decision-theoretic perspective on grounding in conversational systems. *AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems*, Nov 5-7, MA.

Traum, D. 1994. A computational theory of grounding in natural language conversation, Ph.D. diss., Dept. of Computer Science, Rochester.

Traum, D. & Dillenbourg, P. 1996. Miscommunication in multi-modal collaboration. *AAAI Workshop on Detecting, Repairing, And Preventing Human--Machine Miscommunication*, 37-46.

Traum, D. & Dillenbourg, P. 1998. Towards a normative model of grounding in collaboration,. *ESSLLI-98 Workshop on Mutual Knowledge, Common Ground and Public Information*.