MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

From the Programmer's Apprentice to Human-Robot Interaction: Thirty Years of Research on Human-Computer Collaboration

Charles Rich, Candace L. Sidner

TR2006-052 July 2006

Abstract

We summarize the continuous thread of research we have conducted over the past thirty years on human-computer collaboration. This research reflects many of the themes and issues in operation in the greater field of AI over this period, such as knowledge representation and reasoning, planning and intent recognition, learning, and the interplay of human theory and computer engineering.

AAAI Conference on Artificial Intelligence

From the Programmer's Apprentice to Human-Robot Interaction: Thirty Years of Research on Human-Computer Collaboration

Charles Rich and Candace L. Sidner

Mitsubishi Electric Research Laboratories
201 Broadway
Cambridge, MA, 02139, USA
rich,sidner@merl.com

Abstract

We summarize the continuous thread of research we have conducted over the past thirty years on human-computer collaboration. This research reflects many of the themes and issues in operation in the greater field of AI over this period, such as knowledge representation and reasoning, planning and intent recognition, learning, and the interplay of human theory and computer engineering.

Human-Computer Collaboration

Figure 1 illustrates our overall research methodology, which has been to model human-computer collaboration on what is known about human-human collaboration. Furthermore we have focused almost exclusively on the special case of two copresent collaborators, i.e., where each collaborator is able both to communicate with and observe the actions of the other. Examples of such collaborations include two mechanics working on a car engine together or two computer users working on a spreadsheet together. To a first approximation, our approach has been simply to substitute a computer agent for one of the human collaborators, keeping as much else the same as possible.

Due to space limitations, we will *not* attempt to review all research on human-computer collaboration, but limit ourselves to viewing this topic through the lens of our own work and that of our immediate collaborators. Consistent with this, note that bibliography below contains only publications by ourselves and our immediate collaborators.

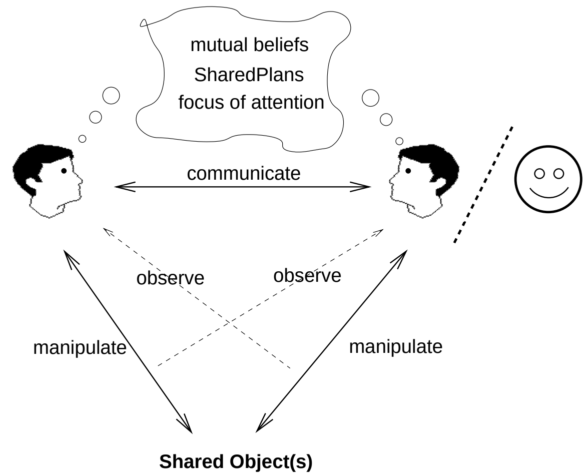

Figure 1: Modeling human-computer collaboration on human-human collaboration.

Chronological Summary

The chronology of our research begins in 1976 with the publication of Rich and Shrobe's joint M.S. thesis on the Programmer's Apprentice [1,3]: "As compared to automatic programming research, the programmer's apprentice emphasizes a cooperative relationship between the computer and the human programmer..." Shortly thereafter, Sidner began work on modeling how natural language is *used* in the context of pairs (and later groups) of people achieving tasks together. Her first paper on this topic dealt with the interpretation of discourse purposes in the Personal Assistant Language Understanding Program [2].

Under the direction of Rich and Shrobe, and later Waters, the Programmer's Apprentice project [4,15,16] lived at the MIT AI Lab from 1976 until Rich and Waters left MIT in 1991. Even though the concept of human-computer collaboration was the bedrock of the project, we never developed a deep theoretical understanding of what collaboration meant. Instead, most of the Programmer's Apprentice research concentrated on how to represent and reason with the shared knowledge necessary for successful human-computer collaboration in the domain of software engineering, including requirements analysis [14], design, and implementation.

In retrospect, the choice of software engineering as a domain, as compared to, for example, medical diagnosis (which was another popular AI application domain at the time) may have been unwise. We were initially attracted by the fact that we already knew a lot about software engineering (as compared to having to spend the equivalent of a year in medical school to learn enough to do research). However, it turned out that the knowledge underlying software engineering is particularly hard to codify, in part because it is difficult to separate from knowledge about the world in which the software is intended to function.

Meanwhile, Grosz and Sidner [5,8,11,12,17,19,21] were delving deeply into the nature of human collaboration, culminating in the SharedPlan theory of collaborative discourse. By 1994, Rich and Sidner [22,23] had begun developing a practical application-independent tool, called Collagen (for *coll*aborative *agen*t), which implemented parts of this theory (see Figure 2). In a sense, Collagen was "the Programmer's Apprentice without the programming."

Collagen continues to evolve [24,26,27,28,32] and has been used to build prototype human-computer collabora-

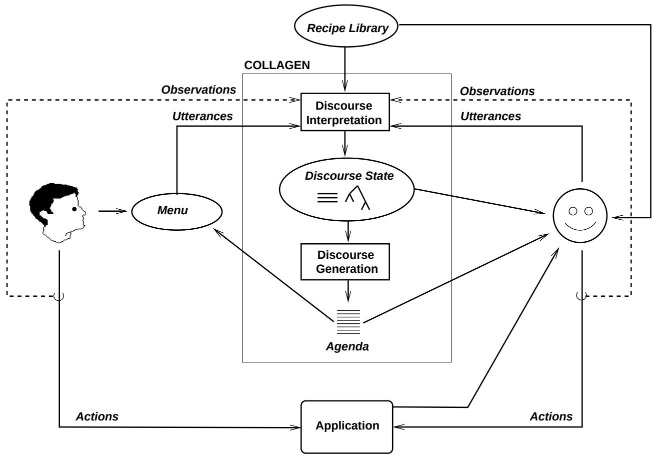

Figure 2: Architecture of Collagen

tive systems, both at MERL and at several other institutions, for a wide range of applications, including air travel planning [28], email [30], gas-turbine operator training [34], programmable thermostat operation [35], power system operation [37], airport landing path planning [38], GUI design [41], and personal video recorder operation [45].

Most recently, we have concentrated on applications of Collagen in the area of feature-rich digitally-enabled home appliances, such as home entertainment systems and programmable combination washer-dryers. In particular, we have developed a new user interface design, called DiamondHelp [50,52,51], which reinforces the collaborative metaphor through the use of a "chat window" between the user and a collaborative agent.

Finally, recognizing the importance of physically embodied, nonverbal behavior [33], such as looking and nodding [42,46,48,49,53], in human collaboration, Sidner has recently embarked on a research program in human-robot interaction to study and apply these phenomena, using hosting activities as a domain.

Knowledge Representation and Reasoning

As in any AI undertaking, the issues of how to represent and reason with the knowledge have been a central concern. For human-computer collaboration, these issues can be broken down along several dimensions. First, there is application-independent versus application-specific knowledge. A major contribution of the theoretical work on discourse has been to show that, given the appropriate abstractions, a significant amount of structure and computation can be captured in an application-independent tool like Collagen.

A second important dimension is computational cost. We are interested in *interactive* collaborative systems, which means that the system needs to respond in a reasonably short period of time. We have therefore been driven to explore hybrid knowledge representations [7,9] which combine expressive general logical formalisms with specialized data structures for fast computation. We are also employing fast, sound, but incomplete inference methods [20].

Collaboration also involves changing your mind and and/or adapting to changes in your collaborator's beliefs. To support this behavior, we make use of dependency-directed inference methods, such as truth maintenance systems [13].

Collaborative Discourse

Collaboration is a process in which two or more participants coordinate their actions toward achieving shared goals. *Discourse* is an extended communication between two or more participants in a shared context, such as a collaboration. From our point of view, discourse and collaboration are virtually coextensive concepts. It is possible to have a discourse without collaboration or a collaboration without discourse, but these are the unusual cases.

Grosz and Sidner [12] developed a tripartite framework for modeling collaborative discourse structure. The first (intentional) component records the beliefs and intentions of the discourse participants regarding the tasks and subtasks to be performed. This component, which came to be known as SharedPlans [17], was further formalized and generalized by Grosz and Kraus [25]. The second (attentional) component captures the changing focus of attention in a discourse using a stack of focus spaces organized around the discourse purposes. The third (linguistic) component consists of the contiguous sequences of utterances, called segments, which contribute to a particular task or subtask.

This model of collaborative discourse has implications for both natural language and speech understanding [45] and generation [47]. Lochbaum [29] developed algorithms for discourse interpretation based on this model, which made it possible to begin implementing Collagen.

More recently, Sidner and colleagues have begun to study the role of nonverbal behavior in collaborative discourse, specifically *engagement*, the process by which participants in an interaction start, maintain, and end their perceived connection to one another in a physical setting. Based on observation of human-human interactions, they have developed a computational model of how nodding and looking behaviors contribute to engagement and tested this model using a physical robot that interacts with a human [42,48,49,53].

Planning and Intent Recognition

A key insight in the Programmer's Apprentice work was the need to represent the programmer's intent at a more abstract level than source code. The Plan Calculus [6] formalism combined concepts from planning and software engineering for this purpose. There was also a significant effort in so-called "program understanding" (also called "reverse engineering") to automatically recover programmer's intent from extant source code [18].

These two themes of planning and intent recognition continue in the Collagen project. Key application-independent services provided by Collagen include: plan-based response generation [44], plan execution monitoring, and limited plan repair and replanning. Furthermore, even though plan recognition is in general NP-complete, Collagen uses a plan recognition algorithm [10,31,39] which is tractable by virtue of exploiting distinguishing properties of the collaborative setting: the focus of attention, the use of partially elaborated hierarchical plans, and the possibility of asking for clarification.

Intelligent Tutoring Systems

Collaboration is a very broad concept which, depending on the relative knowledge and initiative of the participants, spans interactions from helping to teaching; or to put it in human-computer terms, from intelligent assistants to intelligent tutoring systems. Rickel used Collagen to develop PACO (*P*edagogical *A*gent for *CO*llagen) [34,43] for teaching procedural tasks, and as the first step in building a bridge between the intelligent tutoring and the collaborative discourse communities [36].

Learning

Learning is a hallmark of intelligence. The need for many different forms of learning naturally arise in the process of developing human-computer collaborative systems. One obvious case we have pursued is learning hierarchical task models from examples [40]. There is also "learning by being told," which is not as simple as it sounds, since it may involve negotiation about conflicting beliefs [21]. Other forms of learning during collaboration, such adapting to interaction style of the other participant(s), are still open research problems, some being pursued by others.

Chronological Bibliography

[1] C. Rich and H. E. Shrobe. Initial report on a LISP Programmer's Apprentice. Technical Report 354, MIT Artificial Intelligence Lab., December 1976. Master's thesis.

[2] C. L. Sidner. A progress report on the discourse and reference components of PAL. In *Proc. 2nd Nat. Conf. Canadian Soc. Computational Study of Intelligence*, Toronto, Canada, July 1978.

[3] C. Rich and H. E. Shrobe. Initial report on a LISP Programmer's Apprentice. *IEEE Trans. Software Engineering*, 4(6):456–467, November 1978.

[4] C. Rich, H. E. Shrobe, and R. C. Waters. An overview of the Programmer's Apprentice. In *Proc. 6th Int. Joint Conf. Artificial Intelligence*, pages 827–828, Tokyo, Japan, August 1979.

[5] C. L. Sidner and D. J. Israel. Recognizing intended meaning and speaker's plans. In *Proc. 7th Int. Joint Conf. Artificial Intelligence*, Vancouver, British Columbia, Canada, August 1981.

[6] C. Rich. A formal representation for plans in the Programmer's Apprentice. In *Proc. 7th Int. Joint Conf. Artificial Intelligence*, pages 1044–1052, Vancouver, British Columbia, Canada, August 1981.

[7] C. Rich. Knowledge representation languages and predicate calculus: How to have your cake and eat it too. In *Proc. 2nd National Conf. on Artificial Intelligence*, pages 193–196, Pittsburgh, PA, August 1982.

[8] C. L. Sidner. What the speaker means: The recognition of speaker plans in discourse. *J. Computers and Mathematics*, 9(1), 1983.

[9] C. Rich. The layered architecture of a system for reasoning about programs. In *Proc. 9th Int. Joint Conf. Artificial Intelligence*, pages 540–546, Los Angeles, CA, 1985.

[10] C. L. Sidner. Plan parsing for intended response recognition in discourse. *Computational Intelligence*, 1(1), 1985.

[11] B. J. Grosz and C. L. Sidner. Discourse structure and the proper treatment of interruptions. In *Proc. 9th Int. Joint Conf. Artificial Intelligence*, pages 832–839, Los Angeles, CA, August 1985.

[12] B. J. Grosz and C. L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

[13] Y. A. Feldman and C. Rich. Reasoning with simplifying assumptions: A methodology and example. In *Proc. 5th National Conf. on Artificial Intelligence*, pages 2–7, Philadelphia, PA, August 1986.

[14] C. Rich, R. C. Waters, and H. B. Reubenstein. Toward a Requirements Apprentice. In *Proc. 4th Int. Workshop on Software Specification and Design*, pages 79–86, Monterey, CA, April 1987.

[15] C. Rich and R. C. Waters. The Programmer's Apprentice: A research overview. *IEEE Computer*, 21(11):10–25, November 1988.

[16] C. Rich and R. C. Waters. *The Programmer's Apprentice*. Addison-Wesley, Reading, MA and ACM Press, Baltimore, MD, 1990.

[17] B. J. Grosz and C. L. Sidner. Plans for discourse. In P. R. Cohen, J. L. Morgan, and M. E. Pollack, editors, *Intentions and Communication*, pages 417–444. MIT Press, Cambridge, MA, 1990.

[18] C. Rich and L. M. Wills. Recognizing a program's design: A graph-parsing approach. *IEEE Software*, 7(1):82–89, January 1990.

[19] K. E. Lochbaum, B. J. Grosz, and C. L. Sidner. Models of plans to support communication: An initial report. In *Proc. 8th National Conf. on Artificial Intelligence*, pages 485–490, Boston, MA, July 1990.

[20] C. Rich and Y. A. Feldman. Seven layers of knowledge representation and reasoning in support of software development. *IEEE Trans. Software Engineering*, 18(6):451–469, June 1992.

[21] C. L. Sidner. An artificial discourse language for collaborative negotiation. In *Proc. 12th National Conf. on Artificial Intelligence*, pages 814–819, Seattle, WA, August 1994.

[22] C. Rich. Negotiation in collaborative activity: An implementation experiment. *Knowledge-Based Systems*, 7(4):268–270, December 1994.

[23] C. L. Sidner. Negotiation in collaborative activity: A discourse analysis. *Knowledge-Based Systems*, 7(4):265–267, December 1994.

[24] C. Rich and C. Sidner. Adding a collaborative agent to graphical user interfaces. In *Proc. 9th ACM Symp. on User Interface Software and Technology*, pages 21–30, Seattle, WA, November 1996.

[25] B. J. Grosz and S. Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357, October 1996.

[26] C. Rich and C. Sidner. Segmented interaction history in a collaborative interface agent. In *Proc. Int. Conf. on Intelligent User Interfaces*, pages 23–30, Orlando, FL, January 1997.

[27] C. Rich and C. Sidner. Collagen: When agents collaborate with people. In *Proc. 1st Int. Conf. on Autonomous Agents*, pages 284–291, Marina del Rey, CA, February 1997.

[28] C. Rich and C. Sidner. Collagen: A collaboration manager for software interface agents. *User Modeling and User-Adapted Interaction*, 8(3/4):315–350, 1998.

[29] K. E. Lochbaum. A collaborative planning model of intentional structure. *Computational Linguistics*, 24(4):525–572, December 1998.

[30] D. Gruen, C. Sidner, C. Boettner, and C. Rich. A collaborative assistant for email. In *Proc. ACM Conf. on Computer Human Interaction, Extended Abstracts*, pages 196–197, Pittsburgh, PA, May 1999.

[31] N. Lesh, C. Rich, and C. Sidner. Using plan recognition in human-computer collaboration. In *Proc. 7th Int. Conf. on User Modelling*, pages 23–32, Banff, Canada, June 1999.

[32] C. Rich, C. Sidner, and N. Lesh. Collagen: Applying collaborative discourse theory to human-computer interaction. *AI Magazine*, 22(4):15–25, 2001.

[33] J. Cassell, Y. Nakano, T. Bickmore, C. Sidner, and C. Rich. Non-verbal cues for discourse structure. In *Proc. 39th Annual Meeting of the Assoc. for Computational Linguistics*, pages 106–115, Toulouse, France, 2001.

[34] J. Davies, N. Lesh, C. Rich, C. Sidner, A. Gertner, and J. Rickel. Incorporating tutorial strategies into an intelligent assistant. In *Proc. Int. Conf. on Intelligent User Interfaces*, pages 53–56, Santa Fe, NM, January 2001.

[35] E. DeKoven, D. Keyson, and A. Freudenthal. Designing collaboration in consumer products. In *Proc. ACM Conf. on Computer Human Interaction, Extended Abstracts*, pages 195–196, Seattle, WA, March 2001.

[36] J. Rickel, N. Lesh, C. Rich, C. Sidner, and A. Gertner. Building a bridge between intelligent tutoring and collaborative dialogue systems. In *Proc. 10th Int. Conf. on Artificial Intelligence in Education*, pages 592–594, San Antonio, TX, May 2001.

[37] J. Rickel, N. Lesh, C. Rich, C. Sidner, and A. Gertner. Using a model of collaborative dialogue to teach procedural tasks. In *Working Notes of AI-ED Workshop on Tutorial Dialogue Systems*, pages 1–12, San Antonio, TX, May 2001.

[38] B. Cheikes and A. Gertner. Teaching to plan and planning to teach in an embedded training system. In *Proc. 10th Int. Conf. on Artificial Intelligence in Education*, pages 398–409, San Antonio, TX, May 2001.

[39] N. Lesh, C. Rich, and C. Sidner. Collaborating with focused and unfocused users under imperfect communication. In *Proc. 9th Int. Conf. on User Modelling*, pages 64–73, Sonthofen, Germany, July 2001.

[40] A. Garland, K. Ryall, and C. Rich. Learning hierarchical task models by defining and refining examples. In *First Int. Conf. on Knowledge Capture*, Victoria, B.C., Canada, October 2001.

[41] J. Eisenstein and C. Rich. Agents and GUIs from task models. In *Proc. Int. Conf. on Intelligent User Interfaces*, pages 47–54, San Francisco, CA, January 2002.

[42] C. L. Sidner and M. Dzikovska. Hosting activities: Experience with and future directions for a robot agent host. In *Proc. Int. Conf. on Intelligent User Interfaces*, pages 143–150, San Francisco, CA, January 2002.

[43] J. Rickel, N. Lesh, C. Rich, C. Sidner, and A. Gertner. Collaborative discourse theory as a foundation for tutorial dialogue. In *6th Int. Conf. on Intelligent Tutoring Systems*, pages 542–551, Biarritz, France, June 2002.

[44] C. Rich, N. Lesh, J. Rickel, and Garland A. A plug-in architecture for generating collaborative agent responses. In *Proc. 1st Int. J. Conf. on Autonomous Agents and Multiagent Systems*, Bologna, Italy, July 2002.

[45] C. L. Sidner and C. Forlines. Subset languages for conversing with collaborative interface agents. In *Int. Conf. on Spoken Language Processing*, September 2002.

[46] C. Lee, N. Lesh, C. Sidner, L. Morency, A. Kapoor, and T. Darrell. Nodding in conversations with a robot. In *Proc. ACM Int. Conf. on Human Factors in Computing Systems*, April 2004.

[47] D. Devault, C. Rich, and C. L. Sidner. Natural language generation and discourse context: Computing distractor sets from the focus stack. In *17th Int. Florida Artificial Intelligence Research Symp.*, pages 887–892, Miami, FL, May 2004.

[48] C. L. Sidner, C. Lee, C. Kidd, N. Lesh, and C. Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1-2):104–164, 2005.

[49] L.-P. Morency, C. Lee, C. L. Sidner, and T. Darrell. Contextual recognition of head gestures. In *Proc. 7th Int. Conf. on Multimodal Interfaces*, 2005.

[50] C. Rich, C. Sidner, N. Lesh, A. Garland, and S. Booth. Collaborative help for networked home products. In *IEEE Int. Conf. on Consumer Electronics*, Las Vegas, NV, January 2005.

[51] C. Rich, C. Sidner, N. Lesh, A. Garland, S. Booth, and M. Chimani. DiamondHelp: A collaborative task guidance framework for complex devices. In *Proc. 20th National Conf. on Artificial Intelligence*, pages 1700–1702, Pittsburg, PA, July 2005.

[52] C. Rich, C. Sidner, N. Lesh, A. Garland, S. Booth, and M. Chimani. DiamondHelp: A new interaction design for networked home appliances. *Personal and Ubiquitous Computing*, 10(2–3):187–190, 2006.

[53] C. L. Sidner, C. Lee, L.-P. Morency, and C. Forlines. The effect of head-nod recognition in human-robot conversation. In *Proc. ACM Conf. on Human Robot Interaction*, pages 290–296, 2006.