

## Engagement by Looking: Behaviors for Robots When Collaborating with People

Candace L. Sidner, Christopher Lee and Neal Lesh

TR2003-130 September 2003

### Abstract

This paper reports on research on developing the ability for robots to engage with humans in a collaborative conversation for hosting activities. It defines the engagement process in collaborative conversation, and reports on our progress in creating a robot to perform hosting activities. The paper then presents the analysis of a study that tracks the looks between collaborators and discusses rules that will allow a robot to track humans so that engagement is maintained.

*Diabrock 2003*

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.



This paper appears in *DiaBruck: The Proceedings of the Seventh Workshop on the Semantics and Pragmatics of Dialogue*, Kruiff-Korbayova and Kosny (eds.) University of Saarland, pp, 123-130, 2003.

# Engagement By Looking: Behaviors for Robots When Collaborating with People

**Candace L. Sidner**  
Mitsubishi Electric Re-  
search Labs  
201 Broadway  
Cambridge, MA 02139  
sidner@merl.com

**Christopher Lee**  
Mitsubishi Electric Research  
Labs  
201 Broadway  
Cambridge, MA 02139  
lee@merl.com

**Neal Lesh**  
Mitsubishi Electric Re-  
search Labs  
201 Broadway  
Cambridge, MA 02139  
lesh@merl.com

## Abstract

This paper reports on research on developing the ability for robots to engage with humans in a collaborative conversation for hosting activities. It defines the engagement process in collaborative conversation, and reports on our progress in creating a robot to perform hosting activities. The paper then presents the analysis of a study that tracks the looks between collaborators and discusses rules that will allow a robot to track humans so that engagement is maintained.

## 1 Introduction

This paper reports on our research toward developing the ability for robots to participate with humans in a collaborative interaction for hosting activities. Engagement is the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake. Engagement is supported by conversation (that is, spoken linguistic behavior), ability to collaborate on a task (that is, collaborative behavior), and gestural behavior that conveys connection between the participants. While it might seem that conversational utterances alone are enough to convey connectedness (as is the case on the telephone), gestural behavior in face-to-face conversation conveys much about the connection between the participants.

Engagement is a process to further collaborations. It accounts for how to undertake a collaboration, and how to maintain it once it begins. Engagement is the means by which one collabo-

orative partner tells the other that he or she intends to continue the interaction. Engagement is conveyed not only by the collaborator with the speaking turn in the interaction, but is also by the non-speaking collaborator. Since the non-speaker cannot use linguistic devices, gestures are a means to indicate the desire to further or discontinue the collaboration. Grounding (Clark, 1996) is a device that is part of engagement. It is a backward functioning device to indicate that what has just been said has been understood. Successful grounding is evidence that the collaboration will continue, but it is only partial evidence. Grounding failures offer evidence, inclusive at best, that one of the partners may wish to disengage. One challenge for the research on engagement is to understand which gestures serve grounding purposes, which serve conversational devices such as turn taking, and which perform other engagement roles.

Collaborative interactions cover a vast range of activities from call centers to auto repair to physician-patient dialogues. In order to narrow our research efforts, we have focused on hosting activities. Hosting activities are a class of collaborative activity in which an agent provides guidance in the form of information, entertainment, education or other services in the user's environment and may also request that the user undertake actions to support the fulfillment of those services. Hosting activities are situated or embedded activities, because they depend on the surrounding environment as well as the participants involved. They are social activities because, when undertaken by humans, they depend upon the social roles that people play to determine the choice of the next actions, timing of

those actions, and negotiation about the choice of actions. In applying our research, physical robots, serving as guides, replace human hosts in the environment. To do so, our goals include understanding the nature of human-to-human engagement, especially the role of gestures. We then apply our findings to robots interacting with people.

The gestures discussed in this paper generally concern looks at/away from the conversational partner, pointing behaviors, (bodily) addressing the conversational participant and other persons/objects in the environment, all in appropriate synchronization with the conversational, collaborative behavior. Other gestures, especially with the hands and face play a role in human interactions as some researchers (Cassell et al 2000, Pelachaud et al, 1996, among others) are discovering. The paper limits its focus to the types of gesture mentioned above. These engagement gestures are culturally determined, but every culture has some set of behaviors to accomplish the engagement task. These arise from two tasks that participants undertake: the need to pay attention to the environment around them, and the need to convey some of the intentions of the participants via the head, hands and body. Other intentions are conveyed by linguistic means. When collaborators are not face-to-face, they have only linguistic devices, and cultural conventions expressed linguistically, to tell their partner that they wish to dis/continue the interaction, and if proceeding, how to further their joint goals. In face-to-face interaction, gestures can take some of the load of collaborative information. Some gestures serve to convey ongoing engagement, while others, such as pointing, fill in details about collaborative actions and beliefs. The engagement rules explored here include both purposes because in robotic behavior the two purposes are often intertwined.

Not only must the robot produce these gestures, but also it must interpret similar behaviors from its collaborative partner (hereafter CP). Proper gestures by the robot and correct interpretation of human gestures dramatically enhance the success of conversation and collaboration. Inappropriate behaviors can cause humans and robots to misinterpret each other's intentions. For example, a robot might look away for an extended period of time from the human, a

signal to the human that it wishes to disengage from the conversation and could thereby terminate the collaboration unnecessarily. Incorrect recognition of the human's behaviors can lead the robot to press on with a conversation in which the human no longer wants to participate.

While other researchers in robotics are exploring aspects of gesture (for example, Breazeal, 2001, Kanda et al, 2002), none of them have attempted to model human-robot interaction to the degree that involves the numerous aspects of engagement and collaborative conversation that we have set out above. Robotics researchers interested in collaboration and dialogue (Fong et al, 2001) have not based their work on extensive theoretical research on collaboration and conversation, as we will detail later. Our work is also not focused on emotive interactions, in contrast to Breazeal among others. For 2D conversational agents, researchers (notably, Cassell et al, 2000 and Johnson et al, 2000) have begun to explore agents that produce gestures in conversation. This work complements that research while also focusing on the special demands of 3D physical devices.

In this paper we discuss our research agenda for creating a robot with collaborative conversational abilities, including gestural capabilities in the area of hosting activities. We will also discuss the results of a study of human-human hosting and how we are using the results of that study to determine rules and associated algorithms for the engagement process in hosting activities. We will also critique our current engagement rules, and discuss how our study results might improve our robot's future behavior.

## **2 Communicative capabilities for collaborative robots**

To create a robot that can converse, collaborate, and engage with a human interactor, a number of different communicative capabilities must be included in the robot's repertoire. Most of these capabilities are linguistic, but some make use of physical gestures as well. These capabilities are:

- (1) Engagement behaviors: initiate, maintain or disengage in interaction;
- (2) Conversation management: turn taking (Duncan, 1974) interpreting the intentions of the

conversational participants, establishing the relations between intentions and goals of the participants and relating utterances to the attentional state (Grosz and Sidner, 1996) of the conversation.

(3) Collaboration behavior: choosing what to say or do next in the collaboration, to foster the shared collaborative goals of the human and robot, as well as how to interpret the human's contribution (either spoken acts or physical ones) to the collaboration.

Turn taking gestures serve to indicate engagement because the overall choice to take the turn is indicative of continuing the interaction. CPs produce and observe in their partners other types of gestures during the conversation (such as beat gestures, which are used to indicate old and new information (Halliday, 1973, Cassell, 2000)). These types of gestures are significant to robotic participation in conversation because they allow the robot to communicate using the same strategies and techniques that are normal for humans, so that humans can quickly perceive the robot's communication.

We assume that humans do not necessarily turn off their own engagement and conversational capabilities when interacting with robots. While this assumption is a strong one and can be tested with operational robots in collaboration with people, we start with this assumption because many human capabilities are not always consciously under human control. If humans do use their normal engagement and conversational capabilities, then robots must recognize these capabilities. At the same time, robots can themselves use equivalent capabilities to successfully communicate with humans. We hypothesize that such use will make interactions with robots easier and more predictable. Obtaining operational robots that recognize human engagement behaviors and perform them requires that the robot must fuse data gathered from its visual and auditory sensors to determine the human gestures and infer the human intentions conveyed with these gestures. It must also make decisions as it takes part in the collaboration about which intentions it will convey by gesture and which by linguistic means through conversation.

Our engagement model describes an engagement process in three parts, (1) initiating a collaborative interaction with another, (2) maintain-

ing the interaction through conversation, gestures, and, sometimes, physical activities, and (3) disengaging, either by abruptly ending the interaction or by more gradual activities upon completion of the goals of the interaction. The rules of engagement, which operate within this model, provide choices to a decision-making algorithm for our robot about what gestures and utterances to produce.

Our robot, which looks like a penguin, as shown in Figure 1, uses its head, wings and beak for gestures that help manage the conversation and also express engagement with its human interlocutor (3 DOF in head/beak, 2 in wings). The robot can only converse with one person at a time because the collaboration models we use for conversation only posit one partner. However, the robot performs gestures that acknowledge the onlookers to the conversation without their being able to converse. Gaze for our robot is determined by the position of its head, since its eyes do not move. Since our robot cannot turn its whole body, it does not make use of rules we have already created concerning addressing with the body. Because bodily addressing (in US culture) is a strong signal for whom a CP considers the main other CP, change of body position is a significant engagement signal. However, we will be mobilizing our robot in the near future and expect to test these rules following that addition.

To create an architecture for collaborative interactions, we use several different systems, largely developed at MERL. The conversational and collaborative capabilities of our robot are provided by the Collagen<sup>TM</sup> middleware for collaborative agents (Rich et al, 2001, Rich and Sidner, 1998, Lochbaum, 1998) and commercially available speech recognition software (IBM ViaVoice). We use a face detection algorithm (Viola and Jones, 2001), a sound location algorithm, and an object recognition algorithm (Beardsley, 2003) and fuse the sensory data before passing results to the Collagen system. The robot's motor control algorithms use the engagement rule decisions and the conversational state to decide what gestures to perform. Further details about the architecture and current implementation can be found in (Sidner and Lee, 2003).



Figure 1: Mel, the penguin robot

### 3 Current engagement capabilities

The greatest challenge in our work on engagement is determining rules governing the maintenance of engagement. Our first set of rules, which we have tested in scenarios as described in (Sidner and Lee, 2003), are a small and relatively simple set. The test scenarios do not involve pointing to or manipulating objects. Rather they are focused on engagement in simpler conversation. These rules direct the robot to initiate engagement with gaze and conversational greetings. For maintaining engagement, gaze at the speaking CP signals engagement, when speaking, gaze at both the human interlocutor and onlookers maintains engagement while gaze away for the purpose of taking a turn does not signal disengagement. Disengagement from the interaction is understood as occurring when a CP fails to take an expected turn together with loss of the face of the human. When the human stays engaged until the robot has run out of things to say, the robot closes the conversation using known rules of conversational closing (Schegeloff and Sacks, 1973, Luger, 1983).

Though the above list is a fairly small repertoire of engagement behaviors, it was sufficient to test the robot's behavior in a number of scenarios involving a single CP and robot, with and without onlookers. Much of the robot's behavior is quite natural. However, we have observed oddities in its gaze at the end of its turn (for example, it will incorrectly look at an onlooker instead of its CP when ending its turn, which signals that the onlooker is expected to speak) as

well as confusion about where to look when the CP leaves the scene.

These conversations have one drawback: they are of the how-are-you-and-welcome-to-our-lab format. However, our goal is hosting conversations, which involve many more activities. While other researchers have made considerable progress on the navigation involved for a robot to host visitors [e.g. (Burgard et al, 1998)] and gestures needed to begin conversation [e.g. (Bruce et al, 2002)], many aspects of interaction are open for investigation. These include producing extended explanations, pointing at objects, manipulating them (on the part of the humans or robots), moving around in a physical environment to access objects and interacting with them. This extended repertoire of tasks requires many more gestures than our initial set. In addition, some of these gestures needed in hosting would be understood as disengagement cues by our first repertoire (looking away from the human speaker for an extended time is indicative of disengagement). So engagement gestures are sensitive to the conversational and collaborative context of use.

To explore hosting collaborations, we have provided our robot with some additional gestural rules and new recipes for action (in the Collagen framework), so that our penguin robot now undertakes hosting through a demonstration of an invention created at MERL. This hosting activity includes engagement behaviors as well as utterances and physical actions to jointly perform the demo. Mel greets a visitor (other visitors can also be present), convinces the visitor to participate in a demo, and proceeds to show the visitor the invention. Mel points to demo objects (a kind of electronic cup sitting on a table), talks (with speech) the visitor through the use of the cup, asks the visitor questions, and interprets the spoken answers, and includes the onlookers in its comments. The robot also expects the visitor to say and do certain activities, as well as look at objects, and will await or insist on such gestures if they are not performed. The entire interaction lasts about five minutes. Not all of Mel's behaviors in this interaction appear acceptable to us. For example, Mel often looks away for too long (at the cup and table) when explaining them, it (Mel is "it" since it is not human) fails to make sure it's looking at the

visitor when it calls the visitor by name, and it sometimes fails to look for a long enough when it turns to look at objects. To make Mel perform more effectively, as well as to understand how people perform in their interactions, we are investigating gesture in human-human interactions.

#### **4 Evidence for engagement in human behavior**

Much of the available literature on gestures in conversation (e.g. Duncan, 1974, Kendon, 1967) provides a basis for determining what gestures to consider, but does not provide enough detail about how gestures are used to maintain conversational engagement, that is, to signal that the participants are interested in what the other has to say and in keeping the interaction going.

The significance of gestures for human-robot interaction can be understood by considering the choices that the robot has at every point in the conversation for its head movement, its gaze, and its use of pointing. The robot must also determine whether the CP has changed its head position or gaze and what objects the CP points to or manipulates. Head position and gaze are indicators of engagement. Looking at the speaking CP is evidence of engagement, while looking around that room, for more than very brief moments, is evidence of disinterest in the interaction and possibly the intention to disengage. However, looking at objects relevant to the conversation is not evidence of disengagement. Furthermore, the robot needs to know that the visitor has or has not paid attention to what it points at or looks at. If visitor fails to look at what the robot looks at, the visitor might miss something crucial to the interaction.

A simple hypothesis for maintaining engagement (for each listening CP) is: Do what the speaking CP does: look wherever the CP looks, look at him if he looks at you, and look at whatever objects are relevant to the discussion when he does. This simple hypothesis is effective because it assures that the listening CP will have the most information from the speaking CP about the interaction. It will allow the listening CP to be prepared to ground the conversation whenever needed as well. When the robot is the speaking CP, this hypothesis means it will ex-

pect perfect tracking of its looking by the human interlocutor. The hypothesis does not constrain the speaking CP's decision choices for what to look and point at.

Note that there is evidence that the type of utterances that occur in conversation affect the gaze of the non-speaking CP. Nikano (Nikano et al, 2003) provides evidence that in direction giving tasks, the non-speaking CP will gaze more often at the speaking CP when the speaking CP's utterance pairs are assertion followed by elaboration, and more often at a map when the utterance pairs are assertion followed by the next map direction.

To evaluate the simple hypothesis for engagement, we have been analyzing interactions in videotapes of human-human hosting activities, which were recorded in our laboratory. In these interactions, a member of our lab hosted a visitor who was shown various new inventions and computer software systems. The host and visitor were followed by video camera as they experienced a typical tour of our lab and its demos. The host and visitor were not given instructions to do anything except to give/take the lab tour. We have obtained about 3.5 hours of video, with three visitors, each on separate occasions being given a tour by the same host. We have transcribed portions of the video for all the utterances made and the gestures (head, hands, body position, body addressing) that occur during portions of the video. We have not transcribed facial gestures. We report here on our results in observing gestural data (and its corresponding linguistic data) for just over five minutes of one of the host-visitor pairs.

The purpose of the investigated portion of the video is a demonstration of an "Iglassware" cup which P (a male) demos and explains to C (a female). P produces a gesture, with his hands, face, and body, gazes at C and other objects. He also point to the cup, holds it and interacts with a table to which the cup transfers data. He uses his hands to produce iconic and metaphorical gestures (Cassell, 2000), and he uses the cup as a presentation device as well. We do not discuss iconic, metaphorical or presentation gestures, in large part because our robot does not have hands with which to perform similar actions.

We report here on C’s tracking of where P looks (since P speaks the overwhelming majority of the utterances in their interaction). Gaze in this analysis is expressed in terms of head movements (looking). We did not code eye movements due to video quality.

There are 82 occasions on which P changes his gaze by moving his head with respect to C. Seven additional gaze changes occurred that are not counted in this analysis because it is unclear to where P changed his gaze. Of the counted look changes, C tracks 45 of them (55%). The remaining failures to track looks (37, or 45% of all looks) can be subclassed into 3 groups: quick looks, nods (glances followed by gestural or verbal feedback), and uncategorized failures (see Table 1).

These tracking failures indicate that our simple hypothesis for maintaining engagement is incorrect. Of these tracking failures, the *nod failures* can be explained because they occur when P looks at C even though C is looking at something else (usually the cup or the table). In all these instances, P offers an intonation phase, either at his looks or a few words after, to which C nods and often articulates with “Mm-hm,” “Wow” or other phrases to indicate that she is following her conversational partner. In grounding terms [23], P is attempting to ascertain by looking at C that she is following his utterances and actions. When C cannot look, she provides feedback by nods and comments. She is able to do this because of linguistic information from P indicating that her contribution is called for. She grounds P’s comments and thereby indicates that she is still engaged. In the nod cases, she also is not just looking around the room, but paying attention to an object that is highly relevant to the demo. In two instances of nods, P looks away from C to something else. In both cases, C is attending to P, and at his intonation pause, C nods.

	Count	% of tracking failures	% of total failures
Quick looks	11	30	13
Nods	14	38	17
Uncategorized	12	32	15

**Table 1: Failures to Track Changes in Looking**

The quick looks and uncategorized failures represent 62% of the failures and 28% of the look changes in total. Closer study of these cases reveals significant information for our robot vision detection algorithms.

In the *quick look* cases, P looks quickly (including moving his head) at something without C tracking his behavior. In eight instances, P looks to something besides C; in three instances, he looks up to C from the cup or table without her awareness. Is there a reason to think that C is not paying attention or has lost interest? P never stops to check for her feedback in these instances. Has C lost interest or is she merely not required to follow? In all of these instances, the length of the quick look is brief (under 1 second, in most cases under .6 seconds). During these, C is either occupied with something else (looking at something, laughing or nodding) and thus misses the look, or the look occurs without an intonation pause to signal that acknowledgement is expected. In only one instance, does P pause intonationally and look at C. One would expect an acknowledgement here even without tracking P’s looks. However, in that instance, C is distracted by the glass and simply fails to notice the look, which is very brief, only .10 seconds.

Of the *uncategorized failures*, the majority (8 instances) occur when C has other actions or goals to undertake. In addition, all of the uncategorized failures are longer in duration (2 seconds or more). For example, C may be finishing a nod and not be able to track P while she’s nodding. Of the remaining three tracking failures, each occurs for seemingly good reasons to us as observers, but may not be known at the time of occurrence. For example, one failure occurs at the start of the demo when C is looking at the new (to her) object that P displays and does not track P when he looks up at her.

This data clearly indicates that our rules for maintaining engagement must be more complex than the “do whatever the speaking CP does” hypothesis. In fact, tracking is critical to the interaction because it allows the listening CP to observe the speaking CP’s behavior. It is not completely necessary at every moment because quick glances away can be disregarded. Furthermore, the speaking CP can be relied upon to pause for verbal feedback when needed. In

those instances where arguably one should track the speaking CP, failure to do so may lead the speaking CP to pause to wait for return of visual attention and perhaps even to restate an utterance, or alternatively to just go on because the lost information is not critical to the interaction. The data in fact suggest that for our robot engagement rules, tracking the speaking CP in general is the best move, but when another goal interferes, providing verbal feedback, when required, will maintain engagement. Furthermore, the robot as tracker can ignore head movements of brief duration, as these quick looks do not need to be tracked.

We can ask whether the rule of “track whenever possible, but allow other goals to interfere” makes sense, in general terms. Since humans often find themselves collaborating in environments that are not peaceful or may not be benign, lookaways to check up on the world around one are sensible and perhaps necessary. When speaking, lookaways to objects of interest may serve the cognitive function of reminders of how to continue the current utterance. So while tracking serves the very useful function of keeping on top of the other CP’s current behavior, it cannot be performed all of the time. Furthermore as long as the other CP (when speaking) intends to maintain engagement, that individual can be relied upon to provide feedback about what is happening when a CP fails to track.

When the robot is the speaking CP, our data suggest that it would be most natural for the robot to seek acknowledgements from the human, especially when the human is looking at something besides the robot. Of course just when the robot should linguistically seek an acknowledgement remains to be accounted for by a theory of grounding in conversation.

## 5 Future directions

An expanded set of rules for engagement requires a means to evaluate them. We want to use the standard training and testing set paradigm common in computational linguistics and speech processing. However, training and test sets are hard to come by for interaction with robots because one must first have a robot that can perform sufficiently complex interactions to create the sets. Our solution has been to approach

the process with a mix of techniques. We have developed a graphical display of simulated, animated robots running our engagement rules. We plan to observe the interactions in the graphic display for a number of scenarios to check and tune our engagement rules. In addition we are now undertaking an evaluation of the robot’s demonstration of the Iglassware cup with subjects who interact when the robot uses a varied set of gestural tracking rules with each subject group.

## 6 Summary

This paper has discussed the nature of engagement in human-robot interaction, and outlined our methods for investigating rules for engagement for the robot. We report on analysis of human-human look tracking where the humans do not always track the changes in looks by their conversational interlocutors. We conclude that such tracking failures indicate both the default behavior for a robot and when it can fail to track without its human conversational partner inferring that it wishes to disengage from the interaction.

## Acknowledgements

The authors wish to acknowledge the work of Charles Rich on aspects of Collagen critical to this effort.

## References

- Beardsley, P.A. 2003. *Piecode Detection*, Mitsubishi Electric Research Labs TR2003-11, Cambridge, MA, February.
- C. Breazeal. 2001. “Affective interaction between humans and robots,” *Proceedings of the 2001 European Conference on Artificial Life (ECAL2001)*. Prague, Czech Republic.
- A. Bruce, I. Nourbakhsh, R. Simmons. 2002. “The Role of Expressiveness and Attention in Human Robot Interaction,” In *Proceedings of the IEEE International Conference on Robotics and Automation*, Washington DC, May.
- Burgard, W., Cremes, A. B., Fox, D., Haehnel, D., Lakemeyer, G., Schulz, D., Steiner, W. & Thrun, S. 1998. “The Interactive Museum Tour Guide Robot,” *Proceedings of AAAI-98*, 11-18, AAAI Press, Menlo Park, CA.

- J. Cassell. 2000. Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. in *Embodied Conversational Agents*, J. Cassell, J. Sullivan, S. Prevost, and E. Churchill (eds.), Cambridge, MA: MIT Press.
- J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan. 2000. "Human Conversation as a System Framework: Designing Embodied Conversational Agents," in *Embodied Conversational Agents*, J. Cassell, J. Sullivan, S. Prevost, and E. Churchill (eds.), MIT Press, Cambridge, MA.
- H.H. Clark. 1996. *Using Language*, Cambridge University Press, Cambridge.
- S. Duncan. 1974. Some signals and rules for taking speaking turns in conversation. in *Nonverbal Communication*, S. Weitz (ed.), New York: Oxford University Press.
- T. Fong, C. Thorpe, and C. Baur. 2001. Collaboration, Dialogue and Human-Robot Interaction, *10<sup>th</sup> International Symposium of Robotics Research*, Lorne, Victoria, Australia, November.
- B. J. Grosz and C.L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3): 175—204.
- M.A.K. Halliday. 1973. *Explorations in the Functions of Language*, London: Edward Arnold
- W.L. Johnson, J.W. Rickel, and J.C. Lester. 2000. "Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments," *International Journal of Artificial Intelligence in Education*, 11: 47-78.
- T. Kanda, H. Ishiguro, M. Imai, T. Ono, and K. Mase. 2002. "A constructive approach for developing interactive humanoid robots. *Proceedings of IROS 2002*, IEEE Press, NY.
- A. Kendon. 1967. Some functions of gaze direction in social interaction. *Acta Psychologica*, 26: 22-63.
- K.E. Lochbaum. 1998. A Collaborative Planning Model of Intentional Structure. *Computational Linguistics*, 24(4): 525-572.
- H.H. Luger. 1983. "Some Aspects of Ritual Communication," *Journal of Pragmatics*. Vol. 7: 695-711.
- Y. Nikano, G. Reinstein, T. Stocky, J. Cassell. 2003. "Towards a Model of Face-to-Face Grounding," *Proceedings of the 41st ACL meeting*, Sapporo, Japan.
- C. Pelachaud, N. Badler, M. Steedman. 1996. "Generating facial expressions for speech," *Cognitive Science*, 20(1): 1-46.
- C. Rich, C.L. Sidner, and N. Lesh. 2001. "COLLAGEN: Applying Collaborative Discourse Theory to Human-Computer Interaction," *AI Magazine, Special Issue on Intelligent User Interfaces*, AAAI Press, Menlo Park, CA, Vol. 22: 4: 15-25.
- C. Rich and C.L. Sidner. 1998. "COLLAGEN: A Collaboration Manager for Software Interface Agents," *User Modeling and User-Adapted Interaction*, Vol. 8, No. 3/4, 1998, pp. 315-350.
- E. Schegeloff, & H. Sacks. 1973. Opening up closings. *Semiotica*, 7(4): 289-327.
- C.L. Sidner and C. Lee. 2003. *An Architecture for Engagement in Collaborative Conversations between a Robot and Humans*, Mitsubishi Electric Research Labs TR2003-13, Cambridge, MA.
- P. Viola and M. Jones. 2001. Rapid Object Detection Using a Boosted Cascade of Simple Features, *IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, pp. 905-910.