Gestures Reflect Cognitive as Well as Interactional Capacities

Candace L. Sidner and Christopher Lee

Mitsubishi Electric Research Labs
201 Broadway
Cambridge, MA 02139 USA
{sidner, lee}@merl.com

Abstract

We argue for the role of gestures to express intentions related to attention in human-robot interactions.

Gestures In Interaction

Gestures are fundamental to human interaction. When people are face-to-face at near or even far distance, they gesture to one another as a means of communicating their beliefs, intentions and desires. When too far apart or in too noisy an environment to use speech, gestures can suffice, but in most human face-to-face encounters, speech and gesture co-occur. According to the claims of [McNeill, 2002], they are tightly intertwined in human cognition. However cognitively entangled, people gesture freely, and the purposes of those gestures are fundamental our work on cognitive robotics.

Gestures convey intentions, beliefs and desires, that is, information about the individuals who use them. Gestures are made with every part of the body: hands, face, posture of the body, body stance, legs, feet. Facial gestures have been carefully studied to interpret emotion [Ekman,1982], while others have considered the use of facial muscles (such as an eye raise) to convey belief as well as intention [Goodwin, 1986], [Kendon, 1967]. Hands are expressive, providing gestures of the newness information to the conversational interactions (so called beat gestures), as well as iconic, metaphorical and deictic indications, the last used to point to objects in the environment [Cassell, 2002]. Body posture and stance can be used to convey emotion (as can legs and feet), but body posture also signals major boundaries in the units of conversation [Cassell et al, 2001a].

Gestures provide many types of information about the individual to the conversation partner. One additional type of information they provide has been noted among scientists who study human interaction. Gestures convey engagement, that is, the attentiveness of one partner to the other during their interaction. The human-robot interaction team at MERL is focused on the nature of engagement in human-machine interaction.

Engagement is the process by which two (or more) participants establish, maintain, and end their perceived connection to one another. This process includes: making initial contact, negotiating their collaboration during the conversation, assessing the intentions of the other participants in remaining engaged, evaluating whether to stay involved, and ending the interaction [Sidner and Dzikovska, 2002]. In face-to-face interaction, gesture is a significant means of performing this task. Engagement, or demands for engagement from one's partner, can be conveyed by all parts of the body. However, gaze, head movement, arm movements and body stance are the principal means of doing so.

While the above comments concern human to human interaction, robots that participate with human interactors need to understand human gestural behaviors and produce them correctly as well. For example, wandering gaze, and gaze to unrelated items or persons, signal a loss of engagement, and may signal a desire to end the interaction. Arm movements to point at relevant objects not only tell the interactional partner what is of interest, but also indicate where to pay attention in the interaction. They can also be used to get attention when a partner's interest has wandered. The direction of the front of one's body and the trajectory of movement also indicate one's primary focus of attention. When a conversational partner must direct his or her body to something besides a conversational partner, some other means of indicating ongoing engagement must be conveyed instead.

Engaging behaviors in a robotic architecture

Our most recent efforts to capture engagement experiment with a robotic penguin that converses with people and collaboratively presents demoes of itself and of a hardware invention from MERL. We have several major goals for the capabilities of this robot. First and foremost, using the Collagen [Rich et al, 2001] subsystem, we model intentions and mutual beliefs of the robot and his human interactor. Second, the robot must not only interpret engagement behaviors on the part of its human counterpart, it must also produce them in concert with what is transpiring in the interaction. In order to do that, the architecture provides the means to fuse sensory information from the vision and sound systems to provide

to the Collagen component information about the human interactor's doings. Furthermore, the architecture must provide the robot with a way to respond appropriately to human behavior and also initiate appropriate gestural and spoken behavior during the interaction. Maintaining a balance between responding to the human and initiating new behavior is possible because of the way the architecture divides sensory/motor behavior from cognitive behavior.

The architecture of this robot divides its overall behavior into a brain and a body. The brain models the conversation and the overall collaboration that is being pursued, using the Collagen collaboration manager mentioned above. The body gathers sensory data, interprets it and provides it to the brain. It accepts back commands for what to do, and also takes note of the current state of the conversation in order to choose some of its next moves with its sensory devices (see Sidner et al, 2005 for a diagram). The robot uses vision algorithms of [Morency et al, 2002], [Viola and Jones, 2001] to find the people in a room and to track the head movements of one of the people, who is the robot's interactional partner. To determine this partner, the robot listens for a voice, and co-locates that voice with one of the faces it sees in the room. It looks at the person's face as it moves around in front of it. The robot turns away from the CP when it must point to an object in the demo, and it also looks to be certain that the person tracks. The robot also interprets certain of the human's head movements as head nods [Morency et al, 2005]. These movements often accompany verbal comments to acknowledge or accept something the person has said. When the movements occur without verbal input, the human intends to convey the appropriate meaning with just the nod. Finally, wing gestures of the penguin signal (1) new information in its utterances [Cassell et al, 2001] and (2) provide pointing gestures to objects in the environment.

Does this robot's non-verbal gestural behavior have an impact on the human partner? The answer is a qualified yes. In experiments with 37 human participants, all subjects were found to turn their gaze to the robot whenever they took a turn in the conversation, an indication that the robot was real enough to be worthy of engagement [Sidner et al, 2005]. Furthermore, in a comparison between the robot using its full engagement behavior versus one that just stared woodenly straight ahead, participants using the fully active robot looked back at the robot significantly more whenever they were attending to the demonstration in front of them. The participants with the active robot also responded to the robot's change of gaze to the table significantly more than the other participants. These participants unconsciously considered the robot a partner to keep engaged with.

Human interactors with the robot also naturally nodded at the robot. They nodded even when the robot could not interpret their nods, but they nodded significantly more when they were told that the robot understood their nods and also got gestural feedback when they nodded [Sidner et al, 2006].

Research questions

Far more research on engagement gestures for robotics and creating engagement with robots remains to be accomplished. Among the many questions to be considered: How do humans stay engaged when directing most of their attention to a task in the environment? What inferences should a robot draw about the human interactor in such cases? How does the robot decide when/if to signal engagement when it is largely engaged by a task rather than a person? How can robots be built to better integrate language production and engagement behaviors?

References

J. Cassell. “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.), MIT Press, 2000.

J. Cassell, Y.I. Nakano, T.W. Bickmore, C.L. Sidner, and C. Rich. “Non-verbal cues for discourse structure." In Proc. of the 39th Meeting of the Assoc. for Computational Linguistics (ACL 2001), pp. 106-115, July 2001.

P. Ekman. Emotion in the human face. Cambridge University Press, 1982.

D. McNeill. Hand and mind: what gestures reveal about thought. University of Chicago Press, 1992.

C. Goodwin. Gestures as a resource for the organization of mutal orientation. Semiotica, 62(1-2): 29-49. 1986.

A. Kendon. Some functions of gaze direction in social interaction. Acta Psychologica, 26: 22-63, 1967.

L.-P. Morency, A. Rahimi, N. Checka, and T. Darrell. Fast stereo-based head tracking for interactive environment. In Proceedings of the Int. Conference on Automatic Face and Gesture Recognition, pages 375–380, 2002.

Morency, L.-P., Lee, C., Sidner, C., Darrell, T. Contextual recognition of head gestures, Proc. Of the Int. Conference on Multimodal Interfaces (ICMI'05), 2005.

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

C. Sidner, C. Lee, C.D.Kidd, N. Lesh, and C. Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1-2): 140-164,August 2005.

Sidner, C., Lee, C., Morency, C., Forlines, C. The Effect of Head-Nod Recognition in Human-Robot Conversation, Proceedings of the ACM Conference on Human Robot Interaction, pp. 290-296, 2006.

C.L. Sidner and M. Dzikovska. Human-robot interaction: engagement between humans and robots for hosting activities. In Proc. of the IEEE Conference on Multimodal Interfaces (ICMI 2002), pp. 123-128, 2002.

P. Viola and M. Jones. Rapid Object Detection Using a Boosted Cascade of Simple Features. In Proc. of IEEE CVPR 2001, pp. 905-910, 2001.