The Initiation of Engagement by a Humanoid Robot

Candy Sidner
BAE Systems AIT
Burlington, MA 01803
Email: candy.sidner@baesystems.com

Chris Lee
Mitsubishi Electric Research Labs
Cambridge, MA 02139
Email: lee@merl.com

Abstract

Initiation of engagement between humans is a rich and complex process. Providing a humanoid robot with the ability to participate adequately in initiating engagement with a human offers some exciting challenges in designing the robot's behaviors and in designing the evaluation experiments to test initiation.

Introduction

Engagement is the process by which interactors start, maintain and end their perceived connection to each other during an interaction. It combines verbal communication and non-verbal behaviors, all of which support the perception of connectedness between interactors. It is essential to understand the process by which engagement transpires as it is part and parcel of our everyday face-to-face interactions. While the verbal channel provides detailed and rich semantic information as well as social connection, the non-verbal channel can be used to provide information about what has been understood so far, what the interactors are each (or together) attending to, evidence of their waning connectedness, and evidence of their desire to disengage. Furthermore, linguistic and non-verbal gestures of engagement must be coordinated during face-to-face interactions.

The overall problem of how participants come to the attention of one another, how they maintain that attention, how they choose to bring their attention to a close, and how they deal with conflicts in these matters is an open matter of investigation. In this paper, we focus on the matter of how engagement begins, that is how interactors start to indicate their perceived connection and establish the initial connection.

The process of initiating engagement

When we begin a face-to-face interaction with another person, the initiation of the engagement process is not just simply saying "hello." We start it by locating them in the visual scene, catching their attention, sharing looks and glances and uttering some form of greeting. Once we have established our connection, we can proceed in whatever fashion is acceptable for the two partners. If our interlocutor is busy with another person or activity, we have to assess choices for the best way to interrupt them or decide not to do so. Interactions that are not face-to-face deal with much of this by artificial means—a telephone ring or an email message, which our intended receiver can choose to respond to if he or she wishes.

This brief description of the initiation of engagement tells only part of the story. The full story is more variable, based on the location of the intended partner, whether the person is where we expect them to be, and various aspects of our personalities which affect how to catch attention, offer greetings and interrupt each other.

Reproducing Human Initiation of Engagement

If the robot were to closely imitate human initiation of engagement, it would perceive not only the person's face (and hence their identity), but also the body position, general level of activity and the person's general awareness of the robot. The robot would have different choices for deciding *how* to initiate engagement based upon knowledge of the overall situation. Its decisions would be informed by:

- the goal that the robot had for that individual,
- robot's beliefs about the urgency of the current goal,
- whether the robot could see the person clearly enough to identify him/her,
- whether the person was attending to the robot or to another task/person,
- previous encounters with the person (both positive and negative),
- the social roles and relationships between person, and
- the personality of the robot (e.g., polite, aggressive but polite, aggressive and impolite, etc).

To fully reproduce human initiation of engagement a robot would need to use its knowledge to flexibly, and with minimal unexpected intrusion, initiate an interaction with a person. However, even a robot with all the knowledge specified above would still need to make decisions about how to proceed in a particular situation. Its decisions on how to engage the person concern whether to initiate engagement by first performing some physical action (body movement, head movement, non-verbal but audible sound), accompanied or not by an utterance, and determining the human's response to this action. An intended response by a person

(for example, a glance, extended glance, utterance, full head turn to the robot, complete turn away from the robot), or a blank look registering no response, or a combination of these responses would require that the robot assess its next step in initiation. This step could be to fully address the person, to attempt to overcome a negative response (such as, a look away) or to deal with a non-response. Overcoming negative responses and non-responses might take several exchanges with the person, and at some point should the robot fail to engage, it must also weigh the decision to accept this failure.

In our research, we are attempting to accomplish the reproduction of human initiation using a robot in communication with humans. Because the state of technology in vision, language understanding, audition, and robot control falls far short of human capabilities, in many instances, the robot cannot perform as a human would due to lack of some aspect of the knowledge listed above. Two principles hence guide our investigation: (1) Use as many as are reliably available sources of knowledge about the situation and the human counterpart to choose a means of engagement initiation, and (2) with unreliable knowledge, attempt to interrupt a person focused on another task as little as possible, relative to the urgency of the robot's goal and the social roles and relationships between human and robot.

In our efforts, our robot is not only initiating engagement with a person, but finding the person with whom to engage. The "Finding" goal is a goal to find a specific person, and other people in the environment are assumed to be resources that can be used to find the desired person. However, should a "resource" person be needed, they also be found and engaged. Locating the desired person and standing before them is the point at which engagement can take place. At that point the robot must determine how to start an interaction.

The goal of finding a desired person must include the ability to find the person in a particular location (for example, the office they normally work in) as well as another location (for example, a meeting room or another person's office). In addition the robot must be able to respond if it notices the desired person on its way to one of these locations, just as humans typically do.

Performing Engagement with a Robot

While the above description is still a relatively idealized sketch of the engagement between the robot and its human counterpart, the current limits of technology require that existing robots back off from the ideal interaction to ones that are doable and guided by the general principles above.

Our current activity with a mobile robot is to reproduce only part of the type of engagement initiation sketched above. Our robot (Figure 1) is currently able to be sent to find a person in a particular room. Unlike other recent robot efforts (Kruijff *et al.* 2006; Michaud *et al.* 2005), our robot does not discover its surroundings, but uses a map of the environment to plan to reach the room (see Figure 2). It is ready to discover the person along the way, or to request help in locating the intended person from a third person if the person cannot be found in the expected place.

Figure 1: Melvin the penguin robot

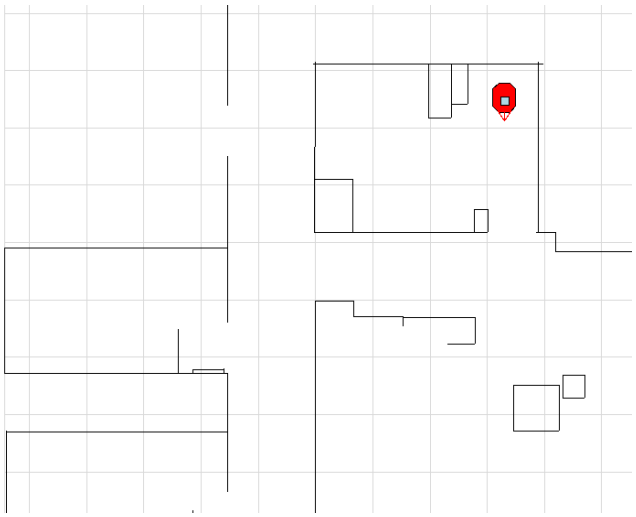

Figure 2: Mel's office environment map

Our robot uses two vision technologies that can track and identify multiple people in a scene. A face detection algorithm (Viola & Jones 2004) locates any individuals in the view as long as they are at least in profile. A face recognition algorithm (Jones & Viola 2003) will recognize faces of any individuals who are within 1.5m of the robot and directly facing it. While face detection and recognition are reliable in good lighting, lighting is often problematic in an office environment (e.g., faces in shadow or in front of bright windows). Hence even if the person we want to is nearby and attending to the robot, the interaction design cannot assume that the robot will necessarily recognize or even see the person. It is thus currently difficult for our robot to judge when a person is attending to the robot or to another task.

Since our robot is not offering objects to individuals, it does not yet have to plan how to approach a person (Hüttenrauch, Green, & Severinson-Eklundh 2005;

Dautenhahn *et al.* 2006). It also does not have sufficient sensors to judge a human form enough to move behind a person when navigating through the environment as the robot in (Dautenhahn *et al.* 2006) does.

We have designed our robot's interactions to engage in three circumstances: human attending, human non-attending and human non-visible. Because it is often difficult for the robot to detect a human and whether the human is attending to the robot, it must have strategies to succeed in each case.

When the human is attending to the robot (which can be sensed by full face or nearly full face presence before the robot), the robot can initiate by looking at the person, offering a greeting and observing the responses by the person.

When the person is non-attending (which the robot can sense only in part using a side face view of the person), our robot currently interrupts the person to ask the location of the desired person ("where is Jones?"). It expects to either be told that the person is the desired person ("I am Jones"), or to be told a location to find him/her ("Jones is in Bill's office"). We believe it is necessary to expand this behavior so that the robot will catch the person's attention based on its social role with the person, its goals, and its own personality traits. Sometimes it will choose to make a noise or move its body to catch the person's attention (Miyauchi *et al.* 2004) as catching attention by movement is less intrusive than a full blown utterance. A polite robot might choose to catch attention, while an aggressive or impolite one or one with an urgent goal might use a full utterance to bring the person's attention to the robot. If the human does not wish to engage, he/she should be able to demand successfully that the robot "go away," "come back later," or say nothing at all.

In cases where no human is visible but there is some expectation that the desired person might yet be nearby (e.g., when the robot is looking for Jones in Jones' office), the robot may announce "I'm looking for Jones," effectively stating its current goal and asking for help with it. This behavior is clearly interruptive, and like the non-attending case, should be expanded to use a means of catching the potential person's attention. However, since no person may actually be in the room, the robot quickly gives up when no response is forthcoming.

Evaluating the engagement initiation process

Demonstrating different types of engagement initiation behavior in an operational setting will provide one set of criteria on which to judge the engagement process of a robot with a person. Ultimately however, we wish to discover the effects of the robot's behavior with individuals. Recent work (Hüttenrauch & Severinson-Eklundh 2003) suggests that many subjects do not respond to a robot asking for help because they are focused on their personal tasks. We wish to understand how they will respond in encounters where the robot is attempting to engage them, but is aware of interruption issues. In evaluations, after-the-fact questionnaires tell us something about a person's experience with the robot. However, we have found in our previous work (Sidner *et al.* 2005; Sidner *et al.* 2006) that video observation provides much more detailed insights into human interaction with robots.

The challenge in our current work is that the robot's environment for usability studies is no longer a single room, as was the case in our previous work, but rather it is an entire floor of our research laboratory! Thus our first challenge is how to instrument that environment so that we can capture what is said and what our "users" do when interacting with the robot. Furthermore, because users may be anywhere within a room when an interaction begins (in addition to the encounters in hallways), we cannot point a camera at a specific location and expect to get the data we want. An alternative is to use human camera staff. We have ruled out this possibility because of the effects of a visibly present set of extra human eyes. Our previous experience with videotaping people in encounters is that they are sensitive to the presence of the camera staff. In this experience, extra people may greatly damp the human individual's response to the robot.

At the present time we are in the discussion stages of how to instrument a floor of a lab to deal with these challenges. We expect that experimentation with the evaluation tools will become part of the evaluation process. The inclusion of environment instrumentation as part of the experimental paradigm seems a suitable problem for social robotics since the field is really in its early stages of development. Part of the solution will probably utilize the onboard sensing and instrumentation of the robot itself.

While our focus concerns the engagement process, the evaluation environment issues we consider above are relevant to any human-robot interaction that is undertaken outside a limited environment because the environment is part of the experimental setting. Ultimately we are trying to build "usability labs" that aren't in a controlled laboratory room, but are out in the world the robot is exploring. In our work we want to have the robot participate in a normal research lab (something like a collection of offices and cubes with people milling around, working and meeting). The more like normal human environments our usability labs are, the more we must take into account the effects of those environments on the testing of human-robot interaction.

Conclusions

Our current work on human-robot initiation of engagement takes into account the differences in robotic capabilities compared to human ones in the capturing of attention, interruption of human activities and social differences in status of the human and robot. Evaluation paradigms for testing the effects of these behaviors include instrumenting large scale normal human environments so that observation of the human-robot interaction behaviors can be studied. While initiation of engagement places certain demands on the paradigm, it shares instrumentation needs with other types of HRI evaluations. Evaluation paradigms for any aspect of interaction that is not in a single room will need to take into account the environments in which the robot is interacting in human-robot evaluations.

References

[Dautenhahn *et al.* 2006] Dautenhahn, K.; Walters, M.; Woods, S.; Koay, K.; Nehaniv, C.; Sisbot, E.; Alami, R.; and Simeon, T. 2006. How may I serve you? A robot companion approaching a seated person in a helping context. In *Proceedings of the 2006 ACM Conference on Human Robot Interaction*, 172–179. New York, NY: ACM Press.

[Hüttenrauch & Severinson-Eklundh 2003] Hüttenrauch, H., and Severinson-Eklundh, K. 2003. To help or not to help a service robot. In *In Proceedings of the 12th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2003)*.

[Hüttenrauch, Green, & Severinson-Eklundh 2005] Hüttenrauch, H.; Green, A.; and Severinson-Eklundh, K. 2005. Report on user study on the role of posture and positioning in HRI. Technical Report IPLab-253, Royal Institute of Technology (KTH).

[Jones & Viola 2003] Jones, M., and Viola, P. 2003. Face recognition using boosted local features. Technical Report TR2003-25, MERL.

[Kruijff *et al.* 2006] Kruijff, G.; Zender, H.; Jensfelt, P.; and Christensen, H. 2006. Clarification dialogues in human-augmented mapping. In *Proceedings of the 2006 ACM Conference on Human Robot Interaction*, 282–289. New York, NY: ACM Press.

[Michaud *et al.* 2005] Michaud, F.; Brosseau, Y.; Ct, C.; Ltourneau, D.; Moisan, P.; Ponchon, A.; Raevsky, C.; Valin, J.-M.; Beaudry, .; and Kabanza, F. 2005. Modularity and integration in the design of a socially interactive robot. In *Proceedings IEEE International Workshop on Robot and Human Interactive Communication*, 172–177.

[Miyauchi *et al.* 2004] Miyauchi, D.; Sakurai, A.; Makamura, A.; and Kuno, Y. 2004. Active eye contact for human-robot communication. In *Proceedings of CHI 2004–Late Breaking Results*, volume CD Disc 2, 1099–1104. ACM Press.

[Sidner *et al.* 2005] Sidner, C. L.; Lee, C.; Kidd, C.; Lesh, N.; and Rich, C. 2005. Explorations in engagement for humans and robots. *Artificial Intelligence* 166(1-2):140–164.

[Sidner *et al.* 2006] Sidner, C. L.; Lee, C.; Morency, L.-P.; and Forlines., C. 2006. The effect of head-nod recognition in human-robot conversation. In *In Proceedings of the ACM Conference on Human Robot Interaction*, 290–296.

[Viola & Jones 2004] Viola, P., and Jones, M. 2004. Robust real-time face detection. *International Journal of Computer Vision* 57(2):137–154.