Emotional Referencing for the Robot Leonardo

Matt Berlin

ABSTRACT

We present a framework for social referencing in an autonomous robotic character. This framework allows the robot to seek out and attend to a human interactor's appraisal of an object when it is uncertain of how to engage with the object for itself. Social referencing creates a new channel of emotional communication between humans and robots, one in which the human plays a central role in shaping the robot's understanding of the objects in its environment. By making robots more attentive to and engaged with humans, we hope to create interactions that are more rewarding and engaging for humans. In this paper, we describe the implementation of the social referencing framework. We describe a simple demonstration built using the new components, and discuss some of the potential applications of the system.

1. INTRODUCTION AND MOTIVATION

I believe that one of the keys for designing systems that are deeply engaging for humans is to create systems that are themselves deeply engaged with humans. Many current robots and other interactive systems fail to be interesting because it's obvious that we don't matter to them - they see us as just another source of input instead of as thinking, acting, social partners. We can make robots more engaging by making humans important to them and to their understanding of the world. An important first step in this process is to design robots that attend to us and learn from us in some of the same ways that we attend to and learn from each other.

To this end, I built preliminary support for emotional referencing into Leonardo, our interactive robot. Emotional referencing is the process by which humans, especially infants, learn an affective response to a novel object or situation by essentially borrowing the affective response of another person. Babies, for example, often learn how to react emotionally to a new object by looking at how their caregiver reacts to that object. Giving Leonardo this ability creates an important channel for emotional communication between the human and the robot. This project fits into a larger, ongoing effort to increase Leonardo's social understanding of objects: where people think they are, what people think they're useful for, how people feel about them, etc.

Adding emotional referencing to Leonardo required the creation and integration of a number of new technologies into the robot's existing behavior and motor control system. First, Leonardo needed to be able to perceive which object in the scene the person was attending to. Second, Leonardo needed to be able to measure the person's emotional reaction to the object - for this project, by classifying their speech as one of a small set of emotional utterances. Third, Leonardo needed to be able to attach this emotional information to his long-term, persistent model of the object. Finally, Leonardo needed to be able to use this emotional information to bias his interactions with the object.

The following section describes these additions to Leonardo's behavior architecture. Section 3 describes a simple demonstration highlighting Leonardo's new emotional referencing abilities. Section 4 describes potential applications and future work involving the new system, including ideas for using the emotional referencing system as a platform for studying deep engagement.

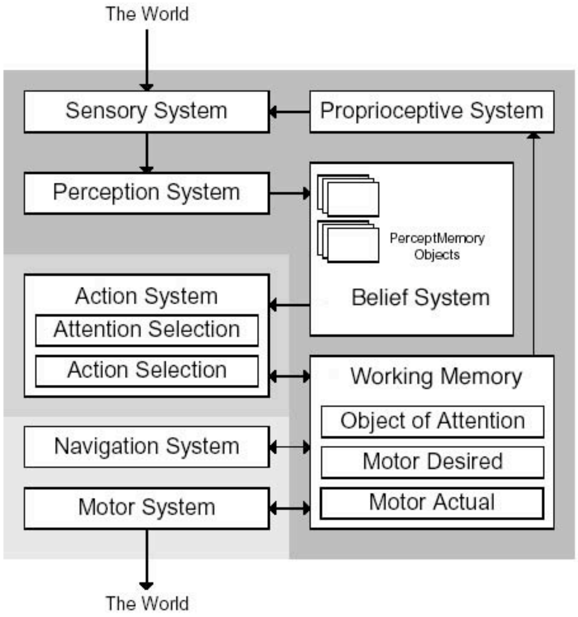

Figure 1: The c5 agent architecture.

2. IMPLEMENTATION

This section describes the additions that were made to Leonardo's behavior architecture to support emotional referencing. I describe improvements to Leonardo's ability to make eye contact as well as progress on a number of new input systems for measuring the human's attentional and emotional state. I describe the creation of a core emotional model for Leonardo, and discuss improvements to Leonardo's emotional signaling abilities.

2.1 Eye Contact

Eye contact is critical for emotional referencing and for human engagement more generally. Leonardo's ability to make eye contact has been hampered by incomplete and inaccurate mappings from his various visual input systems to his local reference frame. For this project, I implemented a new calibration system that significantly improved the accuracy of Leo's spatial understanding of the objects in his environment. This enabled Leo to reason about the spatial relationships between objects, which was critical for tracking the human's object of attention, as described below. This also enabled Leo to look directly at objects, including human faces, and thus significantly increased the frequency with which Leo made eye contact successfully.

I used the new calibration system to build calibration maps for the two main visual input systems used in this project: a face tracking system, and a color-based toy tracking system. To calibrate these systems, a person would stand at a number of different locations in the space in front of Leonardo. The robot would attempt to look either at the person's face or at the brightly colored object that the person carried. The robot would improve his initial tracking of the object's position using feedback from the camera mounted in his right eye. The eye direction would be iteratively adjusted until the tracked object occupied the center of the image produced by this camera.

When this happened, the operator clicked a button, and the calibration system recorded the association between the object's "real" position and its position as reported by the visual input system. Since the human stood at some pre-measured distance from the robot, the "real" position could be calculated by projecting a ray along the direction of the eye out to this distance. The calibration routine continued until a number of these associations between tracked positions and real positions had been recorded. The system then used a singular value decomposition to learn a linear mapping between these tracked inputs and real outputs.

Using this system, I built calibration mappings for the robot's two main visual systems. These calibration mapping were then used online by Leonardo to accurately convert input from his visual system into his local reference frame. From personal experience, I can attest that Leo's improved eye contact and visual attending significantly increased the engagement that I felt while working with him. Studying this effect empirically might be an interesting area for future research.

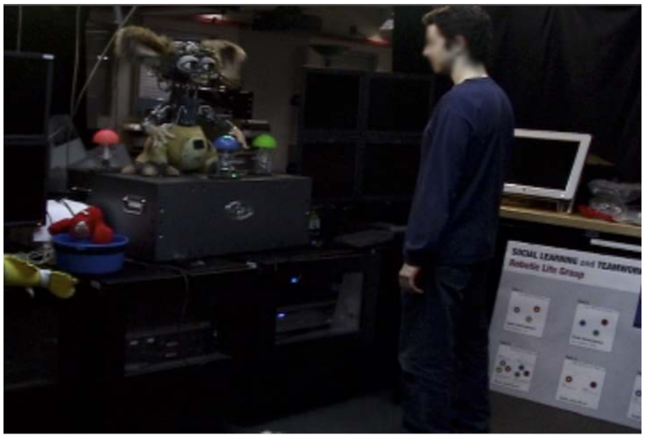

Figure 2: Leo makes eye contact.

2.2 Novel Attentional and Emotional Inputs

I worked on developing a number of new input systems for Leonardo to increase his ability to measure the human's attentional and emotional state. To support social referencing, it is critical that the robot be able to both identify the human's object of attention or referent as well as recognize the human's affect.

2.2.1 Identifying the Referent

I created a new component for Leo's perceptual system that allows him to identify which object in the scene the person is attending to. This new component receives input from a head pose tracker based on the WATSON adaptive tracking system developed by Louis-Philippe Morency and Ali Rahimi. This system uses adaptive view-based appearance models to track the position and orientation (six degrees of freedom) of the closest head in the robot's environment. The head model is built and adjusted online during the tracking process.

To robustly track the human's referent, I implemented a voting mechanism that assigns a score to each of the objects in the robot's environment. An object receives one vote on every timestep that it is looked at (within some margin of tolerance), loses one vote on every timestep that it is not looked at, and loses five votes whenever another object is looked at. The object with the most votes over some threshold is identified as the referent of the communication between the human and the robot.

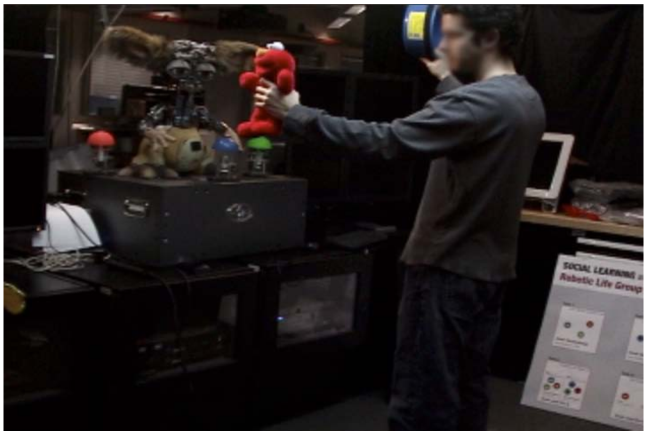

Figure 3: Leo attends to the referent.

2.2.2 Recognizing Affect

I worked on a number of different systems for recognizing the human's affect, but only implemented one system fully within the timeframe of the project. The implemented system uses speech recognition to match the human's speech against a small set of available emotional utterances. The system recognizes a number of different appraisals of the various toy objects that Leonardo can interact with, for example: "Leo, Elmo is your friend" or "The fish is nice" or "The bucket is bad" and so on. The system takes input from Sphinx-4, an open-source, Java-based speech recognition system created by the Sphinx group at Carnegie Mellon University, in collaboration with Sun Microsystems Laboratories, Mitsubishi Electric Research Labs, and Hewlett Packard.

I also worked on systems for recognizing the human's emotional state from their facial expression and vocal intonation. The facial expression recognition system currently takes input from the camera mounted in Leonardo's right eye and uses software developed by the Neven Vision corporation to track some of the human's facial features. Of the features tracked, the eyebrows seem to be the most promising in terms of their reliability and relevance to affective state.

The vocal intonation-based recognition system uses the Praat phonetic analysis toolkit developed by Paul Boersma and David Weenink to track the pitch of recorded speech. Based on my initial work with this system, I think that pitch variability and overall harmonicity may be reliable enough features to generate a useful emotional classification.

Of these various input systems, the speech recognition system is obviously the most impoverished as an affective channel. I believe that continuing to develop one or both of the other input systems will be critical for the success of the project as it progresses. The facial expression recognition system is perhaps the most interesting of these in the context of emotional referencing, since it provides real informational motivation for the robot to actually look back at the human when it encounters a novel object. Looking back at the human would thus be more than just a cosmetic behavior; it would be a critical, motivated component of the interaction. However, facial expression recognition is probably the least reliable of these inputs in an unconstrained interaction. My guess is that intonation-based recognition will probably prove to be the most reliable and generally applicable source of affective information. An interesting experiment might be to selectively turn off one or more of these affective channels in order to determine their relative importance for creating a successful, engaging interaction.

2.3 Core Emotional Model

I developed a core emotional model for Leonardo to support social referencing. The model was designed to enable three important functions. First, it allows Leo to map the human's perceived affective state into his own emotional space. Second, it allows Leo to attach the human's perceived appraisal of the various objects in the scene to his own long-term, persistent memories about these objects. Third, it allows Leo to use these remembered appraisals to bias his emotional signaling and overall behavior towards the objects in the scene.

I created a simple emotional space for Leonardo centered around a valence variable. Leo's valence represents a positive or negative attitude about his environment. Positive appraisal utterances from the human map to a high valence value, and negative appraisal utterances map to a low valence value. In addition to valence, Leo keeps track of his level of familiarity with the various objects he encounters. Leo also keeps track of the relative salience of the objects in the scene, and looks toward the most salient object.

Leo's core appraisal mechanism works as follows. The human's referent receives a high salience value, encouraging Leo to look toward this object. If the referent is unfamiliar, the human's salience increases, causing Leo to look back and forth between the referent and the human. A lack of familiarity with the objects in the scene also causes Leo's valence state to be strongly influenced by the perceived emotional state of the human. If the human seems to be in a positive mood, Leo's valence will increase. If the human seems to be in a negative mood, Leo's valence will decrease. If an object has been the referent for a significant period of time and if Leo's valence is extreme enough, he will attach his current valence state to his long term, persistent memory of the referent object. Thus for novel objects, Leo is biased to take the human's appraisal as his own.

If Leo is attending to a familiar object, his current valence state is influenced by his remembered appraisal of the object. This in turn signals his appraisal back to the human, providing the opportunity for intervention if the human deems his appraisal to be incorrect. Since Leo's appraisals are attached to his long term object memories, they persist even if the object is removed from the scene and brought back much later on. Further, the remembered appraisals can even affect Leo's valence state if the objects are mentioned by the human in their absence. For instance, if the human asks Leo to find the Elmo doll, the remembered appraisal of the doll will influence Leo's emotional state even if the doll cannot be found.

Currently, Leo's valence state is only externally visible in his facial expression. Moving forward, it will be important for Leo's valence to influence his overall behavior more generally. For example, if Leo has a negative association with the Elmo doll, he might refuse to find the doll when asked, rather than simply display his reluctance on his face.

2.4 Emotional Signaling

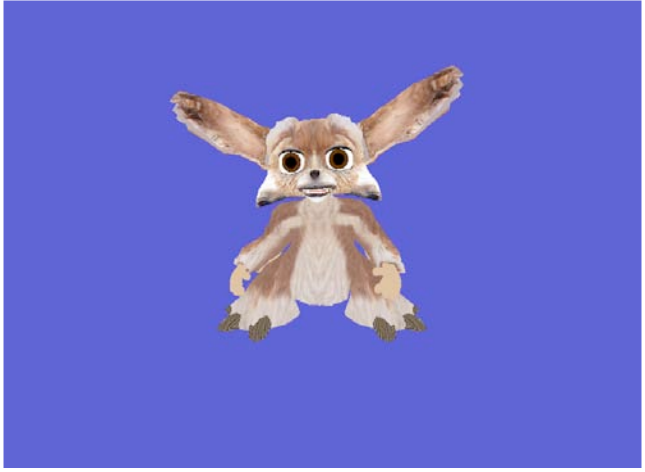

Figure 4: Happy Leo.

I increased Leonardo's ability to convey his emotional state through facial expressions. I added a new facial layer to Leo's motor production system, and populated it with a number of expressive facial poses for use by the robot. Leo expresses his valence state by blending continuously between a "happy" facial pose and a "frustrated" facial pose. As discussed above, it will be important to have Leo's valence influence his behavior more broadly as this project evolves, from his style and quality of motion to his high-level choice of action.

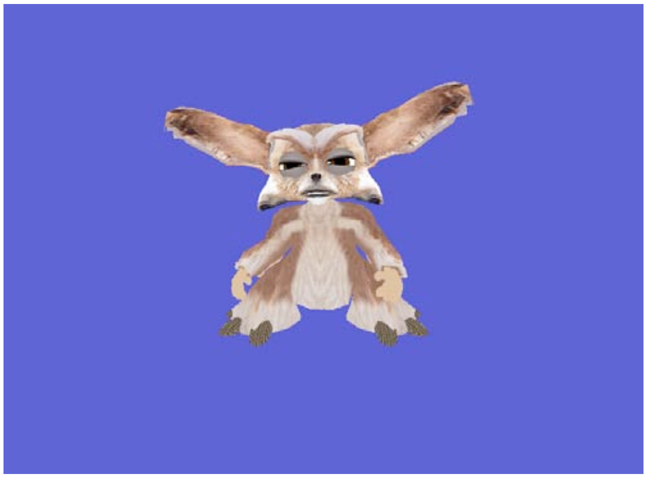

Figure 5: Frustrated Leo.

3. DEMONSTRATION

I integrated the emotional referencing mechanism described above into a simple interactive demonstration involving the Leonardo robot and a single human participant. Leonardo can attend to the human or to any of a number of brightly colored objects in the environment. The human can pick up the objects, move them around, teach Leonardo their names, and influence Leonardo's appraisals of the objects by emotionally expressing at them. Leo's abilities to make eye contact, look at the human's object of attention, and communicate emotionally about the objects with the human all contribute to the engaging quality of the interaction.

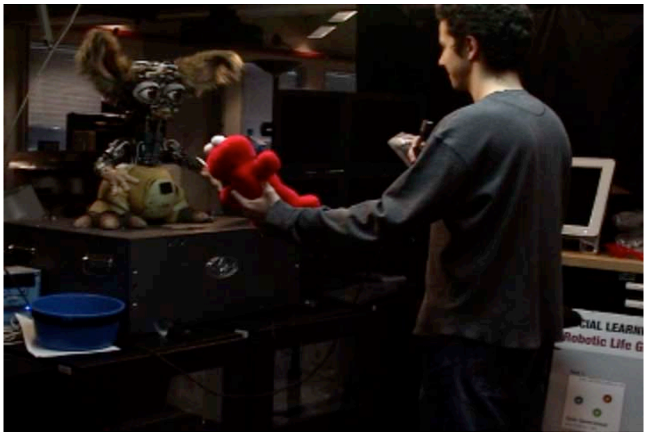

Figure 6: Emotional referencing interaction.

4. APPLICATIONS AND FUTURE WORK

The addition of emotional referencing fits into a larger, ongoing effort to increase Leonardo's social understanding of objects: where people think they are, what people think they're useful for, how people feel about them, etc. We are attempting to build a robot that perceives the world in increasingly social terms, and that increasingly depends on humans for its understanding of its environment. Emotional referencing may play an important role in how the robot learns, and may be useful for establishing turn taking and the self-other divide, and potentially for forming and maintaining trust relationships for learning and teaching.

In the short term, I plan to work on some of the extensions described in the previous sections, such as adding new emotional input channels and expanding the influence of emotion on the robot's visible behavior. One interesting idea, alluded to above, is to use the emotional referencing interaction as a platform for studying the deep engagement process. I am envisioning a simple game based on the demonstration described in the previous section. There would be a number of toys in the space between the participant and Leonardo. Leonardo would shift his attention between the different toys, reacting to them in a way that reflects his appraisal of each object. The person's goal would be to cause Leo to like some of the toys and dislike others of them. In a successful interaction, Leo's assessments of the toys adjust to match the human's assessments.

The success of the interaction could be measured by the participant's performance in the game, as well as by recording video of the interaction and asking the participant to complete a questionnaire. The facial expression and vocal intonation recognition systems might themselves be useful data sources for studying the human's level of engagement. There are many variables that could be manipulated in order to study the engagement process. The accuracy with which the robot makes eye contact and the latency with which he attends to the human's referent could be manipulated to study attentional engagement. Which of the emotion recognition systems the robot uses, as well as their relative importance and reliability, could be manipulated to study emotional engagement. When and how the human's appraisals of the objects influence the robot's memories, and the extent to which those memories affect the robot's behavior, could also be interesting areas for empirical analysis.

All in all, emotional referencing represents an interesting new channel for emotional communication between humans and robots, as well as a promising area for future research. By making robots that are more deeply engaged with humans, that increasingly depend on us and attend to us socially, emotionally, and personally, we may be able to create robots that are themselves more deeply engaging for humans.