# Emotional Robotics: Tug of War

**David G. Cooper** and **Dov Katz** and **Hava T. Siegelmann**
Computer Science Building, 140 Governors Drive
University of Massachusetts, Amherst, MA 01003-9264

### Abstract

Emotional communication skills are dominant in biological systems. Despite the apparent complexity of creating and broadcasting emotional cues, the expression is concise, making them effective and advantageous for multi-agent environments where communication bandwidth is limited and in high demand. However, social robots run the risk of being deceived and used by an opponent using friendly emotional cues. To study this security glitch, we present an interactive environment in which a person can learn the robot's emotional responses through interaction. We then present Tug of War, a game where two people compete for the heart of one robot. The system described is a potential test bed for human-robot interaction, both for engineers, and clinical psychologists.

## Introduction

Emotional communication is a complex interactive process. It may involve multiple agents with different desires and goals. Each agent communicates its emotional state and can deliberately try to manipulate the other's emotional state through interaction. This complex communication process becomes even more intricate when considered in the context of a multi-agent, dynamic and complex environment. Despite its inherent complexity, emotional communication is an efficient way to communicate goals and desires. In this paper, we hypothesize that humans can acquire and adapt to the emotional mechanism that governs the behavior of a robot. Moreover, we demonstrate the crucial role of emotional feedback in facilitating the acquisition of the robot's emotional behavior.

To explore our hypothesis, we have developed an emotional robot with which human participants can interact. Through interactions with the robots, the participants are able to develop a model of its emotional behavior. A reliable model can predict the outcome of future interactions, thus enabling the participant to manipulate the robot to a desired emotional state. The participant's ultimate goal in our experiments is to make the robot happy. Much like preverbal communication with infants, this can be achieved through motion and voice.



Our main contribution is creating a simulated robot that provides a level of expressiveness that is easily understood by the human participants. The robot's behavior must be able to be interpreted within the context of human emotions, in order to enable us to explore this new emotional human-robot interface. To that end, the robot must provide cues that can be understood by the human participants. Moreover, the robot must possess enough sensor capabilities to observe cues generated by the human participants. Based on these cues, both the robot and the human participant can interact. The robot demonstrates its desires, and human participants learn what pleases the robot.

Successful implementation of this robot-human emotional interface would create a platform for testing the degree to which humans can model the emotional state of the robot. This platform will enable us to investigate whether people can successfully attribute cause and effect relations to the behavior of the robot, and then use these relations to manipulate the robot into a state of happiness.

The effectiveness of communication cannot be tested without a task in mind. Only in the presence of some well defined goals can we measure whether an agent is able to accurately and effectively communicate its desires. Therefore, we have added this task dimension to our platform. We chose a simple task that is directly associated with emotional communication: making the robot happy. We describe this task in terms of the popular game *Tug of War* (also known as rope pulling). In *Tug of War*, two teams are competing against each other. A team wins the competition by pulling the other team towards it. Successful teams utilize physical strength, mental strength, and coordination.

We consider two versions of this game for our experiment. The first is a one-player version in which a human tries to convince the robot to move towards it. This can be achieved through emotional communication — by making the robot happy, the robot will be enticed to move towards the human. This version of *Tug of War* enables us to explore the effectiveness and ease of use of our emotional interface. The second version is the standard two-player game. In this version, two competitors try to gain the robot's trust and affection. This version is more intricate than two simultaneous one-player games, as the effect of direct competition between humans adds a new dimension of emotional communication. This version will enable us to explore the reliability of emotional communication for security applications. In these applications, emotional interfaces must be reliable, as they are susceptible to a manipulating enemy that may be able to convince the robot to cooperate with him.

Our implementation relies on cues that are very common in human communication. Running away represents fear. Getting closer signifies trust. Smiling, or putting on an angry face are strong ways to communicate an emotional state. In much the same way, a robot that plays *Tug of War* can elicit and express emotions by moving closer or running away, smiling or frowning.

In this paper, we describe the implementation of the one-player *Tug of War* game. In this version, the goal is making the robot move towards the human participant. This can be achieved in multiple ways, which the participants have to discover during interaction. In some experiments the participants could only see the robot's motion, in others facial expressions were included. To stress the importance of the task, the experiments are scored online. We believe that scores serve as a motivator for the participants to do better. Our experimental results demonstrate that humans can quickly understand the robot's behavior, and are able to interpret it in terms of their own emotions and behavior.

The proposed platform provides an opportunity to conduct clinical psychological research on human participants. The emotional response that the robot generates in humans as part of the competition, as well as the emotional response that the competitors induce on the robot, create a complex emotional interactive environment. Observing this interactive emotional communication will provide an interesting test bed for interdisciplinary research on human and robot emotional communication. We hope that with the development of this test bed, researchers in psychology will be able to provides new insights and develop new models of human-robot emotional interactions.

On the application side, we see believe that adding an emotional aspect to existing Human-Machine interfaces will create a new layer of security. For instance, in the battlefield, robots could choose to cooperate only with people they consider reliable and trustworthy. Databases could be protected by providing information only to the person that convinces the robot it is the rightful owner of that information. Although these applications are very promising, they all can be emotionally manipulated. Our platform will provide a test bed for exploring what measures needs to be taken to overcome this difficulty and guarantee both effective and reliable emotional communication.

In the following sections, we discuss in details the hardware and software components of our platform, as well as the experimental setting and results.

## Emotional Communication Algorithm

### States and Transitions

The emotional algorithm continuously evaluates and acts upon the robot's internal emotional state. The state is represented by an emotional state vector. In the current implementation, the state vector has three states: friend, foe, and self interest. Each state is updated according to the following rule: $s_i \leftarrow s_i + w_i \cdot f_i(I)$, where $i$ is the index of the state vector $s$, $w$ is the weight vector and $f_i$ is function for state $i$ on the input $I$ at the current time step. All weights $w_i$ and states $s_i$ are between 0 and 1, and all functions $f_i$ return a value between 0 and 1. The state vector is always normalized so that the states' values sum to 1.

The expressed emotional state is then based on a which of the state values is above $\frac{1}{2}$. Since the values of the vector are normalized after each update, only one such value can exist. In order to alleviate fast switching between states, a state must be expressed for a minimum of 3 time steps. The expressed states have 3 values for friend and 3 values for foe, and one neutral state. The neutral state must be entered before transitioning between friend and foe states, and a state can change by at most one degree each step. Our implementation is easily extendible to support more dimensions, such as surprise, fear, disgust, and sadness.

The self absorbed state represents times when the robot is incapable of handling inputs. This behavior models periods in which the robot has other needs that have to be fulfilled. These needs include replenishing the battery, some non-interactive tasks, performing off-line learning, and other events that dictate an anti-social behavior. When the robot is in this self-absorbed state, it takes on the neutral expression. The self-absorbed cycle increases at a slow rate for one third of a period, and then decreases at a fast rate for the other two thirds of the period. As the self absorbed value gets larger, it becomes increasingly more difficult to keep the state away from neutral.

### Input Features

The emotional algorithm uses the following 10 visual input features: amount of motion on the screen or face, brightness of screen or face, whether or not the person is facing the robot, the position of the persons face, the amount of motion on the persons face, jittering motion of the face, continuous motion of the face, no motion of the face, and the variance of the face motion. In the experiments described in this paper, we have focused on two of these visual features: whether the person is facing the robot, and the motion variance of the person's face.

The emotional algorithm uses the following 4 audio features: beats per minute (sampled from a second, or using a whole minute), average pitch in the last second (partitioned to high, medium, and low pitch), the mode pitch of the last second, and variance of the pitch. In the experiments described in this paper, we have only used one of these features: the variance of pitch.

We have explored several mappings from inputs to emotional state (the function $f_i$). In the experiments described in this paper, the mapping we use measures whether the person is facing the robot, and the variance in the person's pitch and motion. Facing the agent with which we communicate is an important feature of human communication. The motion variance and the pitch variance measure the jitteriness of the person.

The motion variance of the face is computed over a seven step period. Relatively low variance reflects little change in motion, which makes it easier for the robot to predict what the person is doing. Non-jittery behavior is indicative of comfortable and friendly communication. The pitch variance estimates the change in pitch that is typical of human speech. This was calibrated by the developer's speech and variance is computed over the last second. Regular speech, as opposed to yelling for example, is considered by the robot as a desired method for communication. Finally, we note that the system is designed to allow new input functions to include other inputs such as hand gesture detection or the human's facial expression.

### Output Features

The emotional algorithm is composed of two features: the robot's facial expression, and the robot's motion towards/away from the person. The robot's facial expression and motion is updated after each processing step. Figure 3 shows an example of emotional output with transitions based on emotional state, and Figure 1 are the faces we use. The system's design makes it trivial to add more output features in the future. Possible new outputs can range from adding more facial expressions. to verbal communication.

| Face | How Friendly | Steps Towards Person | Change in Score |
|---|---|---|---|
|  | +3 | +3 | +3 |
|  | +2 | +2 | +2 |
|  | +1 | +1 | +1 |
|  | Neutral | 0 | 0 |
|  | -1 | -1 | 0 |
|  | -2 | -2 | 0 |
|  | -3 | -3 | 0 |

Figure 1: An example of our faces with the emotion, robot behavior, and game score output at each step.

Currently, the output has 7 options for faces, and 7 options for motion, (approach 1-3, withdraw 1-3, and no motion). This creates a wealth of variety for how emotion is expressed in the physical space. Emotion can be expressed by how close the robot is to the person it is interacting with, how fast it is moving towards or away from the person, and whether it is oscillating back and forth. Currently, the gestures of motion are simple. However, extending the motion patterns to express particular emotions would be straight forward.

### Implementation

Our model, which is inspired by (Breazeal 2002; Bar-Cohen & Breazeal 2003), is simple in order to focus purely on the emotional representation. Our model uses sensors, feature extraction, emotional transitions, and emotional expressions, while Breazeal's model has a visual attention system, a cognitive evaluation of the stimuli and drives, a set of higher level drives, an affective evaluation, an affective appraisal, emotion elicitors, emotion activation, behavior, and motor expression. This simplified model separates out the cognitive and reasoning aspects in order to get at the core concept of emotional representation. Rather than the three dimensional space in which emotions lie, we explicitly hold a value for each emotional state. This makes it so that more than one internal emotional state can change at the same time. In this way there can be strict breaks based on one state overwhelming the other states, or smooth transitions over a continuum as the states compete for expression.

Figure 2 shows examples of the emotional states being expressed from screenshots we took of a set of interactions with the simulated robot.

We now describe an example of an interaction between a human and the robot is as follows: The human faces the robot and rocks from left to right slowly for a short time with low variance. This behavior increases the friendliness state, and results in a transition of type **e1** (See Figure 4). The robot smiles and approaches the human (see Figure 3). Next, the person starts to move faster, which displeases the robot as it makes predicting the person's behavior more difficult. In response, the robots changes its facial expression to less friendly values.

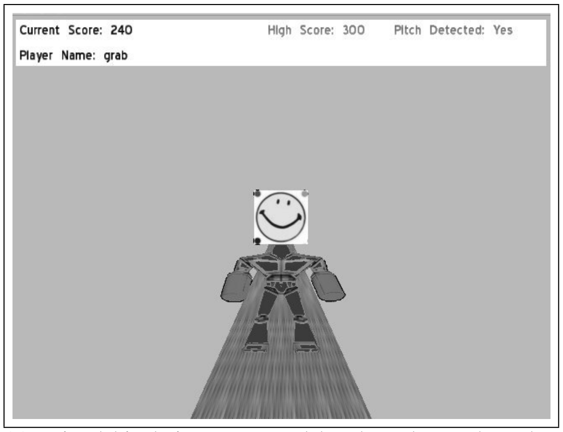
## Experimental Validation

### Experimental Platform

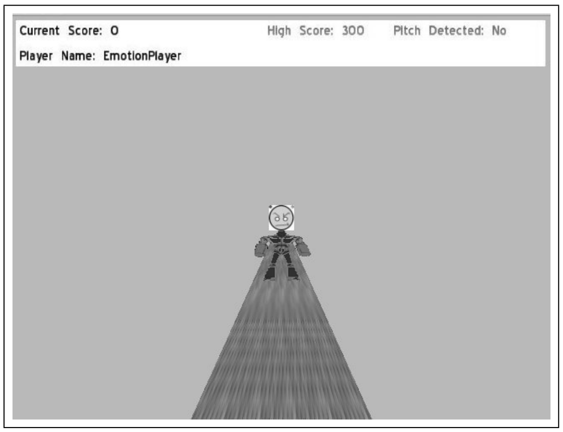
We use a simulated robot and a graphical environment (OGRE 3D) to display robot emotions. When the robot decides to approach the person with whom it interacts, it moves forward in the simulated world. The robot accumulates sensory data from a camera and a microphone, and uses it to decide on transitions between its emotional states. In addition, the robot can express its emotional state using a facial expression.

The emotional architecture is composed of the emotional communication algorithm (described above), integrated with two sensor processing systems, and a 3D simulated robot environment. We use two open source libraries to processes sensor data: OpenCV for vision and MARSYAS for sound processing. With these libraries we were able to quickly produce some basic sensor processing. Consequently, we could focus on the development of the emotional algorithm.

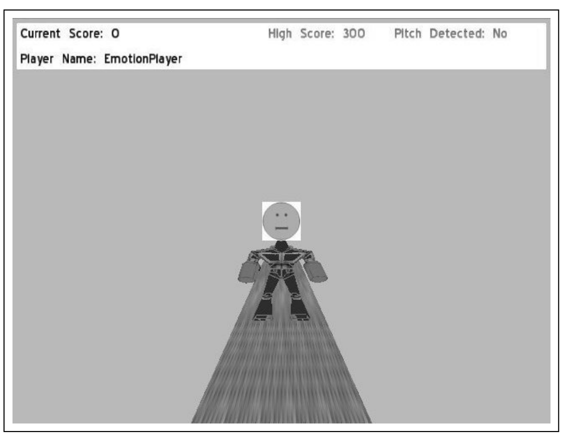
OGRE is a 3 dimensional simulated environment which simulates the motion and cameras of the robot in its environment. In addition, participants can be simulated in the environment by projecting the video from a real camera onto




(a) Friendship being expressed by the robot: The robot approaches and shows a smile as it likes the participant




(b) Dislike being expressed by the robot: The robot withdraws with a scowl as it dislikes the participant.




(c) Indifference being expressed by the robot. The robot stops and has a Neutral face

Figure 2: An example of 3 emotions elicited by a person and expressed by the robot. a) is Friendship, b) is Dislike c) is Indifference

the simulated environment. This allows the robot to move closer and further from the person while the real camera can stay stationary. The software is designed so that the drivers

| | Level 1 | Level 2 | Level 3 |
|---|---|---|---|
| Friend **(e1)** | Small Smile Slow Approach | Med Smile Med Approach | Big smile Fast Approach |
| Foe **(e2)** | Small Scowl Slow Withdraw | Med Scowl Med Withdraw | Big Scowl Fast Withdraw |
| Neutral **(e3)** | Neutral Face Stop | Neutral Face Stop | Neutral Face Stop |

Figure 3: Transitions from emotion expression state to output.

| Weight (friend + foe) | F(I) (friend) | F(I) (foe) | Facing | Motion Var In Range | Pitch Var In Range |
|---|---|---|---|---|---|
| 0.45 | 0.65 | 0 | x | x | x |
| 0.45 | 0.39 | 0 | x | | x |
| 0.45 | 0.39 | 0 | x | x | |
| 0.45 | 0.195 | 0 | | | x |
| 0.45 | 0.195 | 0 | | x | |
| 0.45 | 0 | 0.65 | x | | |
| 0.45 | 0 | 0.65 | | | |

Figure 4: Emotional Input Function based on sensory processing.

for the simulated cameras and robot control can be swapped for the drivers of the real cameras and robot.

The robot can be friend or foe: it can either trust or mistrust the human with whom it interacts. The emotional cues that the robot provides are of two types. The first is a facial expression. The robot has three levels of trust, and three levels of mistrust, and can be indifferent. The second is the distance between the robot and the user. The robot can choose to move closer or further away at three different rates, based on its emotional model.

The sensors we are using are a mono microphone and a web camera. The microphone records at 44.1 kHz. The WebCam has 640 x 480 pixel resolution with 24 bit color and a 10 fps maximum capture rate. The computer that runs the simulation, including sensor and emotional processing, is a 2.33 GHz Intel Core 2 Duo MacBook Pro with 2 gigabytes of RAM.

## Experimental Setup

We have conducted an experiment with 10 participants. Each participant interacted with the robot. Participants experienced two conditions which varied based on the feedback from the robot. In Condition 1 (Face), the robot displays facial expressions and approaches or withdraws from the subject in order to communicate. In addition, a numeric score is displayed. The score is an accumulation of the happy facial expressions. The score increases based on the robot's perception of the friendliness of the human subject. At each time step, the score remained the same or increased by 1, 2, or 3. The score does not change when the robot perceives the human subject as unfriendly, or feels neutral towards it. A score increase of 3 represents the maximal level of friendliness. In condition 2 (No Face), the robot did not change facial expressions. Instead, the robot was constantly displaying a neutral face. The robot displayed a score that was a function of the distance from the subject. At each time step, the score increased by 0 (no motion or moving away from the subject), 1, 2, or 3 (maximum amount of motion per time step towards the subject). The score increased 1 point for each time step that it maintains the minimum allowed distance from the participant.

Participants were randomly divided into two groups of 5. Each participant interacted with the robot in 2 separate sessions. Each session was composed of 2 trials. Both trials were of the same condition, either 2 Face trials or 2 No Face trials. Members of Group A had a face session first and then a No Face session. Members of Group B had a No Face session first, and then a Face session.

The first hypothesis that we evaluate is whether people can improve their interaction with the robot through experience. To verify this hypothesis, we will compare the scores of the second session of each group with the scores of the first session of the other group. For each condition, Face or No Face, if the group that experienced that interaction in the second session (trials 3 and 4) performed better than the other group (which experienced the same condition in their first session (trials 1 and 2), than we can conclude that experience is an important factor in improving the quality of interaction.

The second hypothesis that we evaluate is to what degree facial expressions facilitate the interaction. To verify this hypothesis, we will compare the scores achieved in the first session by each of the groups. If in the first session Group A, which has the Face session first, outperforms Group B, then we have an indication that facial expressions are an important component of communication.

## Experiments

Our first hypothesis, that participants will learn through either interaction condition, is explored in Figures 5 through 7. The higher average score in session 2 of each condition compared with session 1 of the same condition, between groups, indicates that participants do in fact learn the behavior that is needed to make the robot happy in either condition.

Figure 7 shows that there appears to be a trend in the No Face condition. The second trial of each No Face session is an improvement over the first trial. Figure 6 shows that in the Face condition, the opposite effect can be detected. This may be an indication of fatigue in the case of the Face condition. This would suggest boredom if it happened in the second trial of all sessions, but we have only observed this trend for the Face condition. The difference in average values within a session are not statistically significant. In the future, we will use more participants to test whether the reason for the drop can be explained by fatigue.

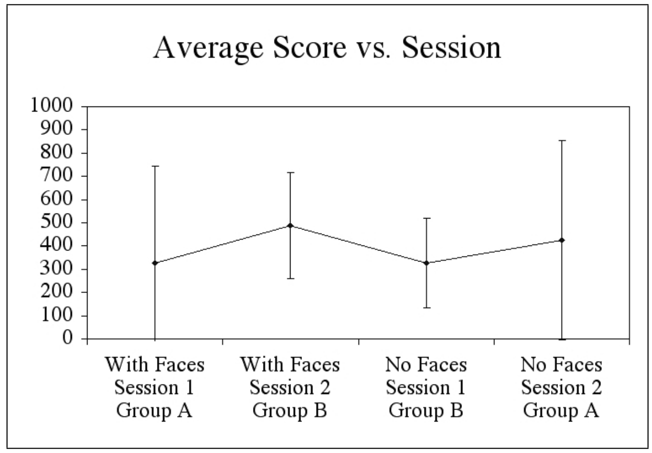



Figure 5: Average score, with standard deviation markers, among each group per session. The values based on the faces condition are on the left and the no faces condition is shown as the right two points. The higher value of Session 2 in each condition indicates that previous exposure to the face condition improves the performance of the no face condition and vice versa. The results are not statistically significant.

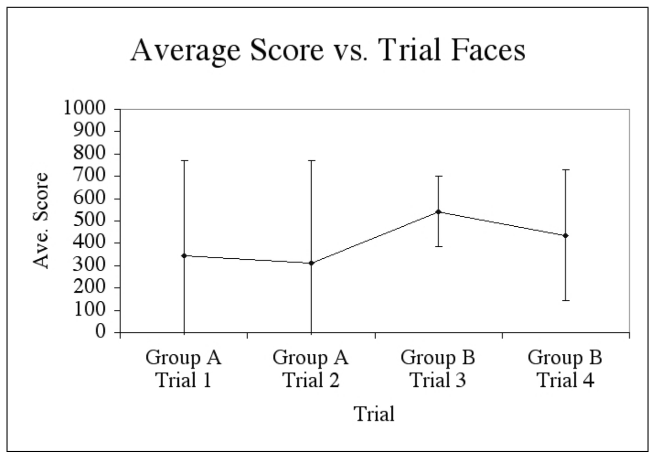



Figure 6: Average score, with standard deviation markers, among the Face trials. The values based on the faces condition appear to increase at later trials, but the second trial of each session decreases. The results are not statistically significant.

To verify our second hypothesis, that facial expressions facilitate the interaction, we look at the values of session 1 in Figure 5. In order for the second hypothesis to be true, the *With Faces Session 1 Group A* value would need to be higher than the *No Faces Session 1 Group B* value. The values are very close and not significantly different, so it is not clear that the facial expressions improved interaction significantly. This result can be interpreted to mean that the display of a numeric score overrides the importance of the robot's facial expression.

Despite the simplicity of our model, the robot expresses

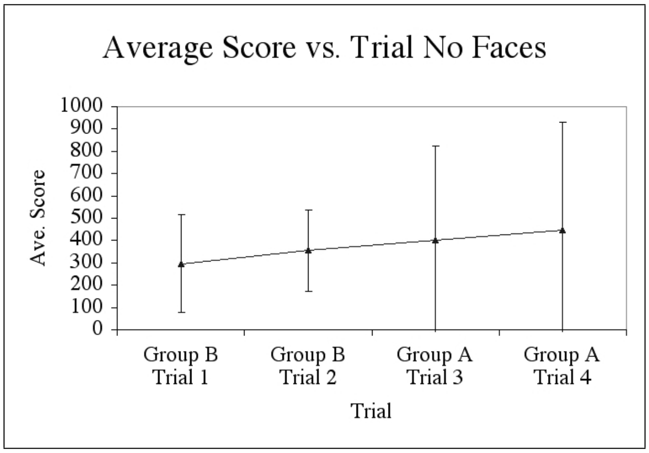



Figure 7: Average score, with standard deviation markers, among each group per Session. The values based on the faces condition appear to increase at later trials, but the second trial of each session decreases. The results are not statistically significant.

enough to let the user know how it feels, which allows the user to continue or change his behavior based on this feedback. Example videos of the robot interaction are available at: *http://binds.cs.umass.edu/EmotionalRobotics.html*

## Conclusion

The experimental results show that the framework we created has promise as an interactive test bed for emotional communication. The emotions elicited by the robot were clear enough for users to know whether or not their actions were pleasing or displeasing for the robot. The simulated environment also allows for repeated testing without having to utilize fragile robotic hardware.

More advanced sensor information could allow for many more options to make the robot happy. For instance if the robot is could do basic speech processing, then there may be words that the robot likes or dislikes. If the robot could do visual shape processing, then certain shapes may be potential inputs. With music processing, certain songs could effect the robot's emotional state.

More advanced expressive output would also contribute to making the robot more believable. For instance, if the robot had arms, it could use them to express emotions by gestures. If it had speakers, it could play different sounds or music. And, with speech software, the robot could say different phrases or change its tone of voice.

We believe that competitive interaction with the robot, such as emotional *Tug of War*, is a useful framework for testing other emotional interactions. With the addition of more advanced input and output capabilities, we believe that this platform could develop even further and provide a more interesting environment for research.

The next phase would be to extend our implementation to create a real *Tug-of-War* game. This will require using two computers, one associated with each contestant. It will also require addressing issues such as synchronizing behaviors based on multiple users, communication over a wireless medium, and doubling the number of sensors used.

Once the system has been tested in a simulated environment, the final step would be putting it onto an actual robot. There are several robotic platforms that could make use of our emotional software (Azad, Asfour, & Dillmann 2007; Brock *et al.* 2005; Brooks *et al.* 2004; Deegan, Thibodeau, & Grupen 2006; Edsinger & Kemp 2006; Katz & Brock 2007; Khatib *et al.* 1999; Neo *et al.* 2006; Nishiwaki *et al.* 2007; Saxena *et al.* 2006; Wimboeck, Ott, & Hirzinger 2007). *UMan*, for example, is a potential future platform. *UMan* is a mobile manipulator, a robot that is both mobile and capable of manipulating its environment. More importantly perhaps, *UMan* has the height of a human, and is able to create an impression on people. *UMan* can support multiple sensors, among which are multiple cameras, force sensors, and laser scanners. For more details, see (Katz *et al.* 2006).

Future work would include discovering ways to characterize the set of human emotions that are easy for the robot to perceive. Once this is done, the input space of the robot could be tuned to human emotion rather than arbitrary sounds or motions. One version of the robot could then have an affinity to happy and angry people. Another version could be attracted to sad and scared people, and put off by happy people.

Finally, we intend to add learning into the emotional agent. One notable characteristic of emotional behavior is the ability to adapt to new circumstances. We would like to create similar behavior in our robot. An agent should learn what is annoying for other people, as well as what is not pleasant for itself. Learning how to achieve goals using emotional reaction can be very beneficial. A robot that can take advantage of emotional communication may be able to communicate more efficiently, and change the state of the world so that it benefits with the robot.

## References

Azad, P.; Asfour, T.; and Dillmann, R. 2007. Toward an Unified Representation for Imitation of Human Motion on Humanoids. In *ICRA*.

Bar-Cohen, Y., and Breazeal, C. 2003. *Biologically Inspired Intelligent Robots*. SPIE Press.

Breazeal, C. 2002. *Designing Sociable Robots*. The MIT Press.

Brock, O.; Fagg, A.; Grupen, R.; Platt, R.; Rosenstein, M.; and Sweeney, J. 2005. A Framework for Learning and Control in Intelligent Humanoid Robots. *International Journal of Humanoid Robotics* 2(3):301–336.

Brooks, R.; Aryananda, L.; Edsinger, A.; Fitzpatrick, P.; Kemp, C.; O'Reilly, U.-M.; Torres-Jara, E.; Varshavskaya, P.; and Weber, J. 2004. Sensing and manipulating built-for-human environments. *International Journal of Humanoid Robotics* 1(1):1–28.

Deegan, P.; Thibodeau, B.; and Grupen, R. 2006. Designing a Self-Stabilizing Robot For Dynamic Mobile Manipulation. In *Robotics: Science and Systems - Workshop on Manipulation for Human Environments*.

Edsinger, A., and Kemp, C. C. 2006. Manipulation in Human Environments. In *IEEE/RSJ International Conference on Humanoid Robotics*.

Katz, D., and Brock, O. 2007. Interactive Perception: Closing the Gap Between Action and Perception. In *ICRA 2007 Workshop: From features to actions - Unifying perspectives in computational and robot vision*.

Katz, D.; Horrell, E.; Yang, Y.; Burns, B.; Buckley, T.; Grishkan, A.; Zhylkovskyy, V.; Brock, O.; and Learned-Miller, E. 2006. The UMass Mobile Manipulator UMan: An Experimental Platform for Autonomous Mobile Manipulation. In *Workshop on Manipulation in Human Environments at Robotics: Science and Systems*.

Khatib, O.; Yokoi, K.; Brock, O.; Chang, K.-S.; and Casal, A. 1999. Robots in Human Environments: Basic Autonomous Capabilities. *International Journal of Robotics Research* 18(7):684–696.

Neo, E. S.; Sakaguchi, T.; Yokoi, K.; Kawai, Y.; and Maruyama, K. 2006. Operating Humanoid Robots in Human Environments. In *Workshop on Manipulation for Human Environments, Robotics: Science and Systems*.

Nishiwaki, K.; Kuffner, J.; Kagami, S.; Inaba, M.; and Inoue, H. 2007. The experimental humanoid robot H7: a research platform for autonomous behaviour. *Philosophical Transactions of the Royal Society* 365:79–108.

Saxena, A.; Driemeyer, J.; Kearns, J.; and Ng, A. Y. 2006. Robotic Grasping of Novel Objects. In *Neural Information Processing Systems*.

Wimboeck, T.; Ott, C.; and Hirzinger, G. 2007. Impedance Behaviors for Two-Handed Manipulation: Design and Experiments. In *ICRA*.