Design of a Social Mobile Robot Using Emotion-Based Decision Mechanisms

Geoffrey A. Hollinger
The Robotics Institute
Carnegie Mellon University
Pittsburgh,PA 15213
gholling@andrew.cmu.edu

Yavor Georgiev, Anthony Manfredi, Bruce A. Maxwell
Department of Engineering
Swarthmore College
Swarthmore, PA 19081
yavor.georgiev@alum.swarthmore.edu, {amanfre1, maxwell}@swarthmore.edu

Zachary A. Pezzementi, Benjamin Mitchell
Department of Computer Science
Johns Hopkins University
Baltimore, MD 21218
{zap, ben}@cs.jhu.edu

***Abstract* – In this paper, we describe a robot that interacts with humans in a crowded conference environment. The robot detects faces, determines the shirt color of onlooking conference attendants, and reacts with a combination of speech, musical, and movement responses. It continuously updates an internal emotional state, modeled realistically after human psychology research. Using empirically-determined mapping functions, the robot's state in the emotion space is translated to a particular set of sound and movement responses. We successfully demonstrate this system at the AAAI '05 Open Interaction Event, showing the potential for emotional modeling to improve human-robot interaction.**

Index Terms – human-robot interaction, robot emotions, face recognition

I. INTRODUCTION

What are emotions? Can robots have emotions? If they could, why would they need them? If they had them, how would we know?

Much philosophical work has been done to try to describe human emotions. David Hume wrote extensively on the passions and their place in the decision making processes. In *A Treatise of Human Nature* [4], he states:

> Since morals, therefore, have an influence on the actions and affections, it follows, that they cannot be derived from reason; and that because reason alone, as we have already proved, can never have any such influence. Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason (3.1.1.6).

Hume sees the emotions (passions) as mechanisms for ethical decision making. According to Hume, moral agents without emotions cannot become excited by ethical decisions and cannot make human-like choices.

Robots without emotions can be likened to decision makers without passions. They decide solely using mechanisms based on reason. Examples of such mechanisms include current implementations of state machines, planning algorithms, and probabilistic decision techniques. Nowhere in these models is there a place for what the robot "feels" like doing. The state transitions are solely rule based.

More recently, Jean-Paul Sartre has looked at emotions and their potential in decision making [8]. In Sartre's view, emotions serve to move agents from one state of mind to another by allowing for quick changes in mental state. According to Sartre, an angry person breaks into an emotional tirade in attempt to change the world by modifying herself. Furthermore, Sartre's interpretation of emotions relates them to the priorities and life-goals of agents. He argues that emotions are, in some manner, intentional and that they serve to guide the agent's actions towards the agent's desired ends in a problematic world.

Relating Sartre's emotional model to robotics, it becomes clear that emotions can serve to guide robotic actions. When determining action-state transitions, for instance, an angry robot could move more quickly from one action to the next in attempt to find a solution. This could potentially be implemented by modifying action-state transition probabilities based on emotion. A happy robot, on the other hand, will likely be warier of changing its actions and would prefer to remain in its current action-state.

In the field of robotics, there has also been work on the place of emotions in robotic decision making. Sloman and Croucher argue that robots must have emotions to prioritize their decision making process [9]. They claim that humans use emotions to determine which life-goals to set above others. For instance, humans often give higher precedence to goals that make them feel happy or fulfilled. Similarly, intelligent robots performing complex tasks must prioritize their activities in some manner. Robot emotions provide an intuitive way of prioritizing actions and determining which task for a robot to perform. Sloman and Croucher's claim also relates back to Sartre and Hume in that it gives emotions a key role in decision making. From these arguments, it seems clear that emotions can provide action guidance for intelligent robots.

Researchers in artificial intelligence have done some previous work in using emotions to guide the actions of artificial agents. For instance, Broekens and DeGroot discuss results from applying emotional models to a Pac-Man playing AI [1], and Gockley et. al present a receptionist robot programmed to display emotions while interacting with humans [3]. Finally, Breazeal and Scassellati use emotional modeling to determine the facial expressions of a robot interacting with people [2]. This research has not, however, used emotional modeling to determine the actions of mobile robots interacting with humans in rich sensory environments. The implementation of robot emotions presented in this paper seeks to specifically explore this area.

II. Problem Description

The robot in this paper was designed for entry into the AAAI Open Interaction Event in 2005. The goal of this event is to entertain conference attendees using robots. The problem is not strictly defined, but robot entries must act autonomously and be able to operate in large crowds.

The robot described in this paper wanders around the conference area using sonar and infrared sensors. Using an onboard camera, it detects faces and determines the presence of onlooking people. When it detects a person, it gives a short verbal and movement response based on the color of the onlooker's shirt and the robot's own current emotional state. Seeing different colored shirts changes the robot's emotional state and alters subsequent responses. When no observers are detected, the robot wanders, and expresses its emotional state through articulate movement and by playing short musical clips.

III. Robot Configuration

A differential-drive RWI Magellan Pro robot was used for the chassis, and the onboard sonar and infrared sensors provided proximity information. A Canon VC-C4 pan-tilt-zoom color camera was mounted near eye level on a metal support on top of the robot chassis. Two generic speakers and a USB Sound Blaster Live! sound card were attached to the robot to provide sound output. This design allowed the robot to perform all of the necessary tasks for the interaction challenge. Fig. 1 shows a picture of the social robot as seen by an onlooker.

Fig. 1 Robot configuration

IV. Principle of Operation

While moving around the conference, the robot uses a state machine to determine which actions it should perform. The robot starts in the Wander state and moves around its environment towards the location of the most open space, using a parameterized proportional controller. In this state, a number of parallel processes are executed:

- *Emotion update*: to account for "boredom"
- *Proportional controller parameter adjustment:* in order to express emotion state through motion
- *Camera color space calibration*: to account for varying lighting conditions

When the robot is in the Wander state and it detects a moving face, it transitions into the Person state. As determined by the emotional model and the person's shirt, it executes a brief movement response, says one line of dialogue, and waits for a given period of time. It then returns to the Wander state and checks for faces. If the onlooker has not moved, the robot says another line of speech as determined by its modified emotional state. Otherwise, it returns to space-following until it detects another onlooker.

V. Implementation Specifics

A. Speech and music

Two different speech segments were pre-recorded for every possible combination of the eight colors the system could differentiate and the twelve potential emotional states. Two short musical clips were selected to represent every emotional state, based on a small-scale study conducted among a group of college students. A piece of music was played and then the subject was asked to describe it in terms of the emotional space used in the system (described exhaustively in Section VI).

B. Movement

The range of movement exhibited by the robot can be characterized in terms of its long-term movement pattern and its reflexes (quick movement responses to interaction with people).

The long-term movement of the robot is based on a free-space following algorithm that uses a proportional controller. A set of parameters is adjusted to modify the movement pattern according to the robot's emotional state:

- k_{trans}, k_{steer} – translational and rotational proportionality constant, modifies the speed of the proportional response
- v_{min}, v_{max} – minimum and maximum movement velocity
- d_{min} – minimum distance to obstacles
- p_{wiggle}, a_{wiggle} – the period and amplitude of the "wiggle," a sideways sinusoidal velocity offset added for more articulate movement

The robot's reflexes are short articulate movement responses (e.g. spinning, swinging, backing up), empirically determined to match particular emotions.

C. Face recognition

To do face detection, OpenCV's [5] object detection function was used. This function is based on the Viola-Jones face detector [10], which was later improved upon by Rainer Lienhart [7]. It uses a large number of simple Haar-like features, trained using a boost algorithm to return a 1 in the presence of a face, and a 0 otherwise. The OpenCV object detector takes a cascade of Haar classifiers specific to the object being detected, such as a frontal face or a profile face, and returns the bounding box if a face is found. An included cascade for frontal faces was used for this system.

Using only the Haar cascade method resulted in a number of false positives from objects such as photographs and posters. To differentiate actual faces from pictures and other "face-like" stationary objects, we added a motion check based on a difference filter. Whenever the Haar detector reports a face, the robot stops, and waits for a set time interval to eliminate any oscillations in the camera boom. Once the camera is perfectly still, the difference operator is executed over a few frames in the bounding box of the face, and the area under it, where the body of the person is supposedly located. If sufficient motion is found (defined by an empirical threshold), the robot transitions from the Wander to the Person state. The motion check coupled with the Haar cascade proved reliable and accurate in all situations where sufficient lighting was present.

D. Color recognition and live camera calibration

After detecting the face bounding box, another box is then defined starting under the person's neck (defined as 0.65 of the height of the face detected), and going down until the bottom of the image or 2 times the height of the face. This box, which should contain the person's shirt, is then used to build a RGB histogram, and its peak is used as the person's shirt color. The color then needs to be classified as one of: red, green, blue, orange, black, white, violet, or yellow. To accomplish this reliably, the RGB color value is then converted to both the HSV color space and the RG/BY opponent color space plus intensity. First, the intensity of the RGB color is checked against a low threshold to determine if the color is dark. Then, high-intensity and low saturation thresholds are used to check if the color is white. If the color does not fall into any of the above two cases, the Euclidean distance is calculated in the HS space against a set of pre-defined colors under neutral lighting.

In order to adjust for variations in color temperature and intensity at the conference venue, we developed a live calibration feature using the PTZ functionality of our camera. A strip of grey photographic cardstock is mounted about 7 cm in front of the camera lens, and, when the camera is tilted down, the strip is visible in the bottom portion of the frame. Thus, by periodically pointing the camera down for about one second, we get a reading for the color of the grey card, which is then matched against a reference estimate to obtain a correction factor. If this grey color measurement is taken under the same neutral lighting conditions as the reference colors described above, the correction factor can be used to obtain an estimate of how the colors would be distorted under the field conditions, which improves the color matching. The approach used during the conference was very similar, but instead of instantaneously determining the correction factor every time the camera was tilted, a weighted average of the previous few tilts was used to filter out random erroneous measurements.

VI. Robot Emotions

For the robot described in this paper, an emotional decision mechanism based on the Mehrabian PAD temperament scale was implemented to determine speech, musical, and movement responses. Similar models have also been used by Broekens and DeGroot [1] and Breazeal and Scassellati [2]. The PAD scale determines emotions using a three-dimensional emotion space with the axes representing pleasure, arousal, and dominance (possible values from -1 and 1). Pleasure represents the overall joy of the agent, arousal its desire to interact with the world, and dominance its feeling of control in the situation. The following emotions exist as points in that space: angry, bored, curious, dignified, elated, hungry, inhibited, loved, puzzled, sleepy, unconcerned, and violent. Violent, for instance, represents a negative pleasure, positive arousal, and positive dominance emotional state. Fig. 2 gives a visualization of the PAD temperament scale as described by Mehrabian [6].

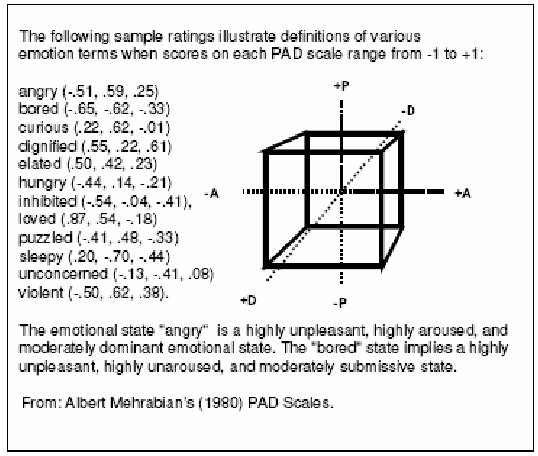

Fig. 2 Mehrabian PAD temperament scale

For the social robot in this paper, interacting with people causes a shift in the robot's PAD temperament, and these shifts affect subsequent dialogue choices. The robot maintains a persistent state in the PAD space, which gets modified when the robot sees different shirt colors. The robot's shirt preferences were arbitrarily selected, although it was attempted to realistically model the effect of seeing a color on a person's state in the PAD space, as described by Valdez and Mehrabian [11]. However, the human results provided very little variation in the PAD space in response to different hues, which would result in passive interaction with onlookers. Therefore, emotional extrema were associated with arbitrary colors as indicated in Table I.

Upon detecting a color, the robot's current state is shifted toward the corresponding emotional extremum, and the step size is randomly selected to be between 5% and 25% of the distance to that extremum. In addition, a random offset ranging from +0.15 to -0.15 is added to each of the three PAD coordinates to produce a more varied response and to ensure that all extrema are reachable. Since the robot is a social robot, it enjoys conversing with people, and it gets lonely when it does not converse with people. Consequently, the robot gains a fixed offset of (0.12, -0.13, 0.24) after conversing for a specified amount of time. Since the robot gets tired of talking after a while, extended conversations cause its arousal to drop. When the robot does not see any humans for a specified period of time, it becomes lonely, and its pleasure, arousal, and dominance all drop by (-0.07, -0.16, -0.09).

TABLE I
COLOR TO EMOTION SPACE MAPPING

Color	Emotion	PAD coordinates
Red	Hostile	(-1, 1, 1)
Green	Docile	(1, -1, -1)
Blue	Relaxed	(1, -1, 1)
Orange	Anxious	(-1, 1, -1)
Black	Bored	(-1, -1, -1)
White	Exuberant	(1, 1, 1)
Violet	Dependent	(1, 1, -1)
Yellow	Disdainful	(-1, -1, 1)

The robot's state in the PAD space directly affects all aspects of its interaction with onlookers. Fig. 3 shows a schematic representation of the mapping between the emotion space and the movement and sound spaces.

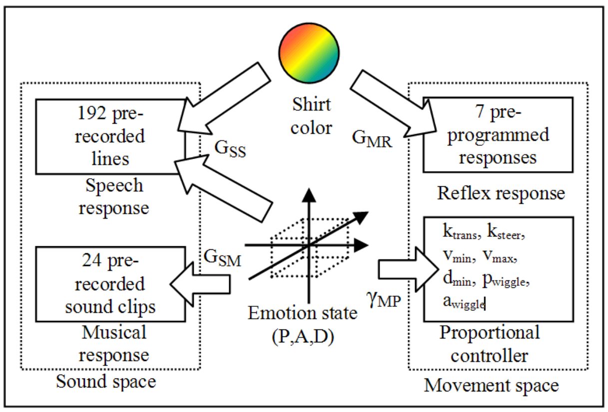

Fig. 3 Emotion space mapping to the movement and sound spaces

We now go on to discuss the four mapping functions used. Let Σ be the set of 12 emotion labels (aliases for points in the PAD space) as listed in Fig. 2. The robot's state is stored as a (P, A, D) triple, which can be associated with an element of Σ by performing a Euclidean distance calculation in the PAD cube (thus effectively subdividing it into 12 volumes, each labeled with a unique element of Σ). Also, let C be the set of 8 colors recognized by the camera, as listed in Table I. Each of G_{MR}, G_{SM}, and G_{SS} takes as its input an element of Σ or C or both and returns one or more output values.

G_{MR} is a function from C to R where R is a set of 7 short articulate reflex responses – e.g. spinning, swinging, backing up – that demonstrate a particular emotion. In our implementation, a value was defined for all elements of C except 'blue'.

G_{SS} is a function from (C, Σ, b) to S where S is a set of 192 pre-recorded verbal comments, and b is a randomly-generated binary variable. S contains two values for each combination of C and Σ, and the variable b determines which one gets selected (192 = 8 colors * 12 emotions * 2 responses for each combination). This way, if a person is interacting with the robot, he or she will hear one of two verbal responses about his or her shirt, presuming the robot remains in the same emotional state. Example members of S for the same color-emotion combination (black, angry) are "Black shirt… Are you a Goth?" and "Black shirt!? Somebody better call the fashion police."

G_{SM} is a function from (Σ, b) to M where M is a set of 24 pre-selected music clips, with two matching each emotional label. Again, b is a binary variable that randomly selects between one of the two clips available for each emotion. Examples of the musical clips selected include Rob Zombie's "Dragula" and John Williams' "Imperial March," which both correspond to the emotional label 'hostile'.

γ_{MP} is the only continuous function used in the model, and it is defined from (P, A, D) to (k_{trans}, k_{steer}, v_{min}, v_{max}, d_{min}, p_{wiggle}, a_{wiggle}). Our implementation used the following mappings, with k_{trans} and k_{steer} being dimensionless quantities, v_{min} and v_{max} measured in m/s, d_{min} and a_{wiggle} measured in m, and p_{wiggle} measured in s:

$$
\begin{aligned}
k_{trans} &= 0.2 + 0.4(A/2+0.5) & d_{min} &= 0.2 + 0.3\,(0.5 - D/2)\\
k_{steer} &= 0.2 + 0.4(A/2+0.5) & p_{wiggle} &= 10\\
v_{min} &= 0.4 + 0.4(D/2+0.5) & a_{wiggle} &= 1.5(P/2+0.5)\\
v_{max} &= 0.4 + 0.4(D/2+0.5)
\end{aligned}
$$

The combined effect of these four functions in response to the robot's inherent state and an input stimulus (color) defines the full range of its behavior. This framework can be extended to a variety of input sensors (with sound being an obvious candidate), and a variety of expression spaces (including a CG or mechanical face).

VII. RESULTS AND TESTING

The robot was tested in Hicks Hall at Swarthmore College to determine if the face recognition and emotional modeling were correctly implemented. The robot performed favorably while interacting with students and professors. The Haar cascade face detector, after being combined with motion detection, worked well at separating humans from the rest of the environment. While there were occasional false positives, the overall performance of the face detector was very reasonable. Most importantly, onlookers did not see the robot's behavior as anomalous. Fig. 4 shows the robot interacting with an onlooker.

Fig. 4 Social robot interacting with onlooker at AAAI '05

Humans interacting with the robot were often curious about the decision mechanism behind the replies. When they were told that emotional modeling was behind it, they became even more interested in the robot, but eventually the diversity of the replies was insufficient to hold their attention. These preliminary observations show that humans are often fascinated by the prospect of robot emotions, and this fascination leads them to interact with the robot.

Quantitative results of onlookers interacting with the robot were gathered at the AAAI Open Interaction Challenge in July 2005. The robot wandered around the conference area and interacted with any person who would acknowledge it. Fig. 5 and Table II present the average interaction times and number of onlookers for various emotional states. For these results, the emotional states were collapsed into three general categories: happy, sad, and angry. In general, the robot spent equal amounts of time in each state. These preliminary results show that more humans interacted with sad and happy robots for longer periods of time while tending to avoid angry robots. Furthermore, onlookers often correctly perceived the emotional state of the robot as angry, sad, or happy. This verifies that the actions of the robot correctly conveyed the emotional stage and shows that emotional modeling on a mobile robot can be effective at modifying onlooker interaction time.

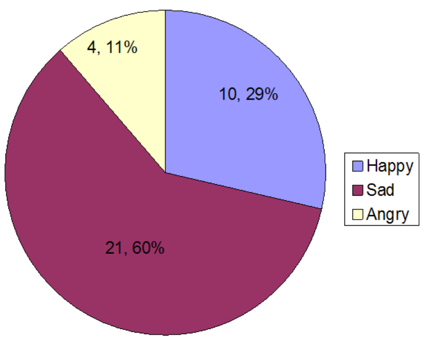

Fig. 5: Number of onlookers for various emotional states

TABLE II
RESULTS OF HUMAN INTERACTION WITH AN EMOTIONAL ROBOT

Emotional State	Total Onlookers	Average Interaction Time (seconds)
Happy	10	46
Sad	21	44
Angry	4	34

VIII. CONCLUSIONS AND FUTURE RESEARCH

In conclusion, this paper has shown that simple emotional models can be effective at holding human attention. AAAI attendants often remained interested in the robot for over a minute at a busy event. The introduction of emotion-based decision mechanisms allowed onlookers to humanize the robot and describe its responses using words like happy, sad, and angry. Furthermore, the tendency of onlookers to avoid angry robots shows that emotional state can be used to affect interaction time. These results show a clear success of emotional modeling, and they demonstrate the potential for mapping PAD emotional values to speech, movement, and music.

For future research, new emotional decision mechanisms should be explored to help in human-robot interaction tasks. For instance, the transitions between the Wander and Person states in this paper were not controlled by the emotional model. However, a slight modification would allow an angry robot to avoid people and a bored robot to find faces in places where there might not be any. These actions would help to humanize the robot and make interacting with it more attractive. Additionally, continuous mappings of PAD values to movement and music could be explored. For instance, human reaction to beats in music and movement patterns could be examined to determine a tighter correspondence between PAD space and robot actions. Furthermore, a large-scale test of interaction times should be conducted. The results in this paper focus on a single conference, and more data would likely yield further insight into how emotional state affects interaction time.

With respect to the emotional model itself, a more sophisticated approach should be explored beyond the Mehrabian PAD temperament model. Future robots could learn emotions through interaction with humans and act in such a way as to mimic them. Since the Mehrabian model is based on Freudian psychology, it brings with it considerable assumptions about human psychology. A learning approach based on neurobiology would likely provide a more compelling emotional model for human-robot interaction.

Finally, the emotional mechanisms in this paper should be applied to a full-scale humanoid robot. This would provide a more informative test platform to examine how people humanize emotional robots, but it requires the formulation of more sophisticated control laws than those developed here. Breazeal and Scassallati present promising work in this direction [2].

The overall goal of this research has been to add an emotional mechanism to the standard reason-guided decision mechanisms in mobile robotics. While there is certainly still a long way to go, as robots gain human-like emotions, they will surely move closer to humans.

ACKNOWLEDGMENT

Thanks to the Engineering Department at Swarthmore College and to the organizers of the AAAI 2005 Robot Exhibition.

REFERENCES

[1] Broekens, J. and D. DeGroot. “Scalable and Flexible Appraisal Models for Virtual Agents,” CGAIDE, 2004.
[2] Breazeal, C. and B. Scassellati, “How to build robots that make friends and influence people,” Proc. Int’l Conf. on Intelligent Robots and Systems, 1999.
[3] Gockley, R. et al. “Designing Robots for Long-Term Social Interaction,” IROS, 2005.
[4] Hume, David. *A Treatise of Human Nature*. New York: Hafner Press, 1948.
[5] Intel Corporation. “OpenCV: Open Source Computer Vision Library” http://www.intel.com/research/mrl/research/opencv/
[6] Mehrabian, Albert. *Basic Dimensions for a General Psychological Theory*. Cambridge: OG&H Publishers, 1980.
[7] Lienhart, R. and J. Maydt. “An Extended Set of Haar-Like Features for Rapid Object Detection,” Proc. IEEE Int’l Conf. Image Processing, vol. 1, pp. 900-903, 2002.

[8] Sartre, John-Paul. *The Emotions: Outline of a Theory*. New York: Citadel Press, 1993.
[9] Sloman, A. and M. Croucher. "Why Robots will have Emotions," IJCAI 1981.
[10] Viola, P. and M. Jones. "Rapid Object Detection using a Boosted Cascade of Simple Features," CVPR, 2001.
[11] Valdez, P. and A. Mehrabian. "Effects of color on emotions." *Journal of Experimental Psychology: General*, 123, 394-409. (1994).