2005

Affective Interaction between Humans and Robots

Cynthia Breazeal[1]

MIT Artificial Intelligence Lab, 200 Technology Square, room 933
Cambridge, MA 02139 USA
`cynthia@ai.mit.edu`
`http://www.ai.mit.edu/projects/kismet`

Abstract. This paper explores the role of emotive responses in communicative behavior between robots and humans. Done properly, affective communciation should be natural and intuitive for people to understand. This implies that the robot's emotive behavior should be life-like. The ability to establish and maintain a rich affective dynamic with people has placed important constraints on our robotic implementation. We present our framework, discuss how these constraints have been addressed, and demonstrate the robot's ability to engage naive human subjects in a compelling and expressive manner.

1 Introduction

Motivated by applications such as robotic pets for children or robotic nursemaids for the elderly, rich affective interchanges will become increasingly important as robots begin to enter long-term relationships with people. The majority of social robotics work took inspiration from ants, termites, fish, and other species that exist in anonymous socities. More recently there has been a shift to taking inspiration from species that live in individualized societies, such as primates, dolphins, and humans [1]. In a similar spirit, this work examines human-robot interaction. Whereas past work in robotics and animated life-like characters has explored the role of computational models of emotions in decision making and learning [2, 3], this paper focuses on the role of emotions in interacting with people on an affective level. Heavily inspired by the study of emotions and expressive behavior in living systems, our approach is designed to support a rich and tightly coupled dynamic between robot and human, where each responds contigently to the other on an affective level. This property is often overlooked, but is critical for establishing a compelling social interaction with humans. It also places important constraints on the implementation of the `emotion` and `expression` systems. We have implemented and evaluated our work on a highly expressive anthropomorphic face robot called Kismet. Human subjects interact with Kismet in the spirit of a human caregiver, robot infant scenario.

2 A Functional and Evolutionary View of Emotions

Emotions are an important motivator for complex organisms. They seem to be centrally involved in determining the behavioral reaction to environmental (often social) and internal events of major significance for the needs and goals of a creature [4]. Several theorists argue that a few select emotions are *basic* or *primary* — they are endowed by evolution because of their proven ability to facilitate adaptive responses to the vast array of demands and opportunities a creature faces in its daily life. The emotions of anger, disgust, fear, joy, sorrow, and surprise are often supported as being basic from evolutionary, developmental, and cross-cultural studies [5]. Each basic emotion is posited to serve a particular function (often biological or social), arising in particular contexts (eliciting conditions), to prepare and motivate a creature to respond in adaptive ways. The orchestration of each emotive response represents a generalized solution for coping with the demands of the original eliciting event. Plutchik (1991) calls this stabilizing feedback process *behavioral homeostasis*. Through this process, emotions establish a desired relation between the organism and the environment — pulling toward certain stimuli and events and pushing away from others. Much of the relational activity can be social in nature, motivating proximity seeking, social avoidance, chasing off offenders, etc.

The expressive characteristics of emotion in voice, face, gesture, and posture serve an important function in communicating emotional state to others. This benefits people in two ways: first, by communicating feelings to others, and second, by influencing others' behavior. For instance, the crying of an infant has a powerful mobilizing influence in calling forth nurturing behaviors of adults. Emotive signaling functions were selected for during the course of evolution because of their communicative efficacy. For members of a social species, the outcome of a particular act usually depends partly on the reactions of the significant others in the encounter. The projection of how the others will react to these different possible courses of action largely determines the creature's behavioral choice. The signaling of emotion communicates the creature's evaluative reaction to a stimulus event (or act) and thus narrows the possible range of behavioral intentions that are likely to be inferred by observers.

3 Design of the Emotion System

The organization and operation of Kismet's *emotion system* is strongly inspired by various theories of emotions in humans and animals. Kismet's `emotions`[1] are idealized models of basic emotions, where each serves a particular function (often social), each arises in a particular context, and each motivates Kismet to respond in an adaptive and expressive manner. Taken together, these emotive responses form a flexible system that mediates between both environmental and internal stimulation to elicit an adaptive behavioral response that serves either

---

[1] As a convention, I will use the boldface to distinguish parts of the architecture of this particular system from the general uses of those words.

social or self-maintenance functions. Summarizing these ideas, an "emotional" reaction for Kismet consists of:

- A precipitating event
- An affective appraisal of that event
- A characteristic expression (face, voice, posture)
- Action tendencies that motivate a behavioral response

Antecedent conditions	Emotion	Behavior	Function
delay, difficulty in achieving goal of adaptive behavior	anger, frustration	complain	show displeasure to caregiver to modify his/her behavior
presence of an undesired stimulus	disgust	withdraw	signal rejection of presented stimulus to caregiver
presence of a threatening, overwhelming stimulus	fear, distress	escape	move away from a potentially dangerous stimuli
prolonged presence of a desired stimulus	calm	engage	continued interaction with a desired stimulus
success in achieving goal of active behavior, or praise	joy	display pleasure	reallocate resources to the next relevant behavior, (eventually to reinforce behavior)
prolonged absence of a desired stimulus, or prohibition	sorrow	display sorrow	evoke sympathy and attention from caregiver, (eventually to discourage behavior)
a sudden, close stimulus	surprise	startle response	alert
appearance of a desired stimulus	interest	orient	attend to new, salient object
need of an absent and desired stimulus	boredom	seek	explore environment for desired stimulus

Table 1. Summary of the antecedents and behavioral responses that comprise Kismet's emotive responses.

Table 1 summarizes under what conditions certain emotive responses arise, and what function they serve the robot. This table is derived from the evolutionary, cross-species, and social functions hypothesized by Plutchik (1991). The table includes the six primary emotions proposed by Ekman (1982) along with three arousal states (boredom, interest, and calm). By adapting these ideas to Kismet, the robot's emotional responses mirror those of biological systems and therefore should seem plausible and readily understandable to people. Figure 1 presents the implementation of the `fear` emotive response to illustrate the relation between the eliciting condition(s), appraisal, action tendency, behavioral response, and observable expression.

For Kismet, some of these responses serve a purely communicative function. The expression on the robot's face is a social signal to the human caregiver, who responds in a way to further promote the robot's "well-being." For instance, the robot exhibits sadness upon the prolonged absence of a desired stimulus. This may occur if Kismet has not been engaged with a toy for a long time. The sorrowful expression is intended to elicit attentive acts from the human caregiver. Another class of affective responses relates to behavioral performance. For instance, a successfully accomplished goal is reflected by a smile on the robot's face, whereas delayed progress is reflected by a frustrated expression. Exploratory responses include visual search for desired stimulus and/or maintaining visual engagement of a desired stimulus. Kismet currently has several protective responses, the strongest of which is to close its eyes and turn away from threatening or overwhelming stimuli. Many of these emotive responses serve a regulatory function. They bias the robot's behavior to bring it into contact with desired stimuli (orientation or exploration), or to avoid poor quality or dangerous stimuli (protection or rejection). Taken as a whole, these affective responses encourage the human to treat Kismet as a socially aware creature and to establish meaningful communication with it.

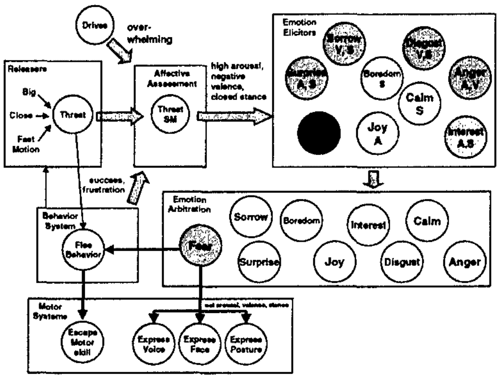


Fig. 1. The implementation of the `fear` emotion. The releaser for `threat` is passed to the affective assessment phase where it is tagged with high arousal, negative valence, and closed stance values. This affective information is then filtered by the corresponding elicitor of each `emotion` process. Darker shading corresponds to a higher activation level. The `fear` process becomes active, causing a fearful expression and evoking `anescape` response.

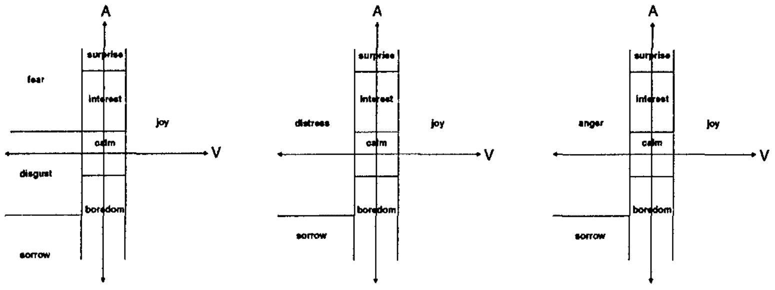
Emotive Releasers The input to the emotion system originates from the high-level perceptual system, where each percept is fed into an associated *releaser* process. Each releaser can be thought of as a simple "cognitive" assessment that combines lower-level perceptual features into behaviorally significant perceptual categories. There are many different kinds of releasers defined for Kismet, each hand-crafted, and each combining different contributions from a variety of factors. These factors include the robot's homeostatic state, its current affective state, the active behavior, and the perceptual state (for details, please refer to

[6]). Hence, each releaser is evaluated with respect to the robot's "well-being" and its goals. If the conditions specified by that releaser hold, then its output is passed to the affective appraisal stage where it can influence the emotion system.

Affective Appraisal Within the appraisal phase, each releaser is appraised in affective terms where the incoming perceptual, behavioral, or motivational information is "tagged" with affective information. There are three classes of tags used to affectively characterize a given releaser. Each tag has an associated intensity that scales its contribution to the overall affective state. The *arousal* tag, A, specifies how arousing this factor is to the emotional system. It very roughly corresponds to the activity of the autonomic nervous system. Positive values correspond to a high arousal stimulus whereas negative values correspond to a low arousal stimulus. The *valence* tag, V, specifies how favorable or unfavorable this percept is to the emotional system. Positive values correspond to a pleasant stimulus whereas negative values correspond to an unpleasant stimulus. The *stance* tag, S, specifies how approachable the percept is. Positive values correspond to advance whereas negative values correspond to retreat. There are four types of appraisals considered:

- *Intensity*: The intensity of the stimulus generally maps to arousal. For instance, threatening or very intense stimuli are tagged with high arousal.
- *Relevance*: The relevance of the stimulus (whether it addresses the current goals of the robot) influences valence and stance. For instance, stimuli that are relevant are "desirable" and are tagged with positive valence and approaching stance.
- *Intrinsic Pleasantness*: Some stimuli are hardwired to influence the robot's affective state in a specific manner. For instance, praising speech is tagged with positive valence and slightly high arousal [6].
- *Goal Directedness*: Each behavior specifies a goal, i.e., a particular relation the robot wants to maintain with the environment. Success in achieving a goal promotes `joy` and is tagged with positive valence. Prolonged delay in achieving a goal results in `frustration` and is tagged with negative valence and withdrawn stance.

Emotion Elicitors This tagging process converts the myriad of factors into a common currency that can be combined to determine the net affective state. For Kismet, the $[A, V, S]$ trio is the currency the emotion system uses to determine which emotional response should be active. All somatically marked inputs are passed to the *emotion elicitor* stage. Each `emotion` process has as elicitor associated with it that filters each of the incoming $[A, V, S]$ contributions. Only those contributions that satisfy the $[A, V, S]$ criteria for that `emotion` process are allowed to contribute to its activation. Figure 2 summarizes how $[A, V, S]$ values map onto each `emotion` process. This filtering is done independently for each type of affective tag. For instance, a valence contribution with a large negative value will not only contribute to the `sad` process, but to the `fear`, `distress`, `anger`, and `disgust` processes as well. Given all these factors, each elicitor computes its average $[A, V, S]$ from all the individual arousal, valence, and stance values that pass through its filter.



Closed Stance   Neutral Stance   Open Stance

Fig. 2. Mapping of arousal, valence, and stance dimensions, [A, V, S], to `emotions`. This figure shows three 2-D slices through this 3-D space.

Given the net $[A, V, S]$ of an elicitor, the activation level is computed next. Intuitively, the activation level for an elicitor corresponds to how "deeply" the point specified by the net $[A, V, S]$ lies within the arousal, valence, and stance boundaries that define the corresponding `emotion` region shown in figure 2. This value is scaled with respect to the size of the region so as to not favor the activation of some processes over others in the arbitration phase. The contribution of each dimension to each elicitor is computed individually. If any one of the dimensions is not represented, then the activation level is set to zero. Otherwise, the A, V, and S contributions are summed together to arrive at the activation level of the elicitor. This activation level is passed on to the corresponding `emotion` process in the arbitration phase.

Emotion Activation and Arbitration Numerically, the activation level $A_{emotion}$ of each `emotion` process can range between $[0, A^{max}_{emotion}]$ where $A^{max}_{emotion}$ is an integer value determined empirically. Although these processes are always active, their intensity must exceed a threshold level before they are expressed externally. The activation of each process is computed by the equation:

$$A_{emotion} = \sum(E_{emotion} + B_{emotion} + P_{emotion}) - \delta_t$$

where $E_{emotion}$ is the activation level of its affiliated elicitor process, $B_{emotion}$ is a DC bias that can be used to make some `emotion` processes easier to activate than others. $P_{emotion}$ adds a level of persistence to the active emotion. This introduces a form of inertia so that different `emotion` processes don't rapidly switch back and forth. Finally, δ_t is a decay term that restores an `emotion` to its bias value once the `emotion` becomes active. Hence, the `emotions` have an intense activation period followed by decay to a baseline intensity on the order of a few seconds.

Next, the `emotion` processes compete for control in a winner-take-all arbitration scheme based on their activation level. Each emotive response becomes active under a different environmental (or internal) situation, and each motivates a different observable response in behavior and expression. In a process of behavioral homeostasis as proposed by Plutchik (1991), the emotive response maintains activity through feedback until the correct relation of robot to environment is established.

4 Emotive Expression

Concurrently, the net $[A, V, S]$ of the active `emotion` process is sent to the expressive components of the motor system, causing a distinct facial expression and body posture to be exhibited. The strength of the facial expression reflects the level of activation of the `emotion`.

There are two threshold levels for each `emotion` process: one for expression and one for behavioral response. The expression threshold is lower than the behavior threshold. This allows the facial expression to lead the behavioral response. This enhances the readability and interpretation of the robot's behavior for the human observer. For instance, if the caregiver shakes a toy in a threatening manner near the robot's face, Kismet will first exhibit a fearful expression and then activate the `escape` response. By staging the response in this manner, the caregiver gets immediate expressive feedback that she is frightening the robot. If this was not the intent, then the caregiver has an intuitive understanding of why the robot is frightened and modifies behavior accordingly. The facial expression also sets up the human's expectation of what behavior will soon follow. As a result, the caregiver not only sees what the robot is doing, but has an understanding of why.

Psychologists such as Smith & Scott (1997) posit that facial expressions have a systematic, coherent, and meaningful structure that can be mapped to affective dimensions. It follows that some of the individual features of facial expression have inherent signal value. For instance, raised brows convey attention in both fear as and surprise. This promotes a signaling system that is robust, flexible, and resilient [7]. It allows for the mixing of these components to convey a wide range of affective messages, instead of being restricted to a fixed facial configuration for each emotion. This variation allows fine-tuning of the expression, as features can be emphasized, de-emphasized, added, or omitted as appropriate.

In keeping with this theory, Kismet's facial expressions are generated using an interpolation-based technique over a three-dimensional *affect space* — the same three $[A, V, S]$ attributes used to affectively assess the robot's siutation (see figure 3). The computed net affective state occupies a single point in this space,

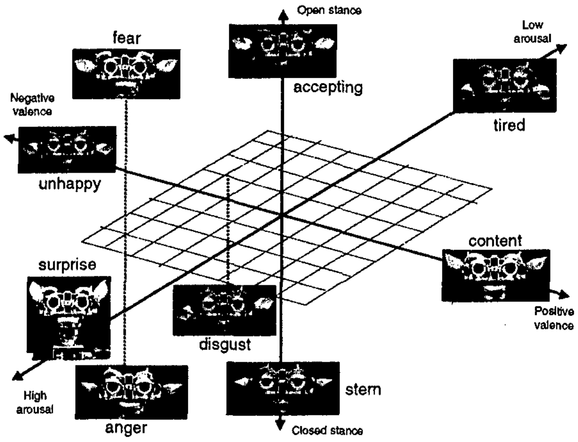


Fig. 3. This diagram illustrates where the basis postures are located in affect space.

moving along a trajectory as the robot's affective state changes. The procedure runs in real-time, which is critical for social interaction. There are nine *basis* (or *prototype*) postures that collectively span this space of emotive expressions. The basis set of facial postures has been designed so that a specific location in affect space specifies the relative contributions of the prototype postures to produce a net facial expression that faithfully corresponds to the active emotion. With this scheme, Kismet displays expressions that intuitively map to the emotions of anger, disgust, fear, happiness, sorrow, and surprise, and many more. Different levels of arousal can be expressed as well from interest, to calm, to weariness. A similar scheme is used to control affective shifts in body posture.

There are several advantages to generating the robot's facial expression from this affect space. First, this technique allows the robot's facial expression to reflect the nuance of the underlying assessment. Even through there is a discrete number of emotion processes, the expressive behavior spans a continuous space. Second, it lends clarity to the facial expression since the robot can only be in a single affective state at a time (by our choice) and hence can only express a single state at a time. Third, the robot's internal dynamics are designed to promote smooth trajectories through affect space. This gives the observer a lot of information about how the robot's affective state is changing, which makes the robot's facial behavior more interesting. Furthermore, by having the face mirror this trajectory, the observer has immediate feedback as to how their behavior is influencing the robot's internal state. For instance, if the robot has a distressed expression upon its face, it may prompt the observer to speak in a soothing manner to Kismet. The soothing speech is assimilated into the emotion system where it causes a smooth decrease in the arousal dimension and a push toward slightly positive valence. Thus, as the person speaks in a comforting manner, it is possible to witness a smooth transition to a subdued expression.

5 Dynamic Affective Exchanges with Humans

To explore the affective coupling between Kismet and human subjects, we carried out the following experiment. Five female subjects, ranging from 23 to 54 years old, were asked to either praise, scold, alert, or soothe Kismet through tone of voice, and to signal when they felt that Kismet understood them. None had interacted with Kismet previously. All sessions were recorded on video for further evaluations. For each trial, we recorded the number of utterances spoken to the robot, Kismet's expressive feedback cues, subject's responses and comments, as well as changes in tone of voice, if any. Kismet's ability to recognize these affective intents has been reported in [6]. To enduce a change in "emotional" state and to express this state to a human, the output of the affective intent recognzier is fed through the `emotion` and `expression` systems as presented in this paper.

Recorded events show that subjects in the study made ready use of Kismet's expressive feedback to assess when the robot "understood" them. The subjects varied in their sensitivity to the robot's expressive feedback, but all used facial expression and/or body posture to determine when the utterance had been properly communicated to the robot. All subjects would reiterate their vocalizations with variations about a theme until they observed the appropriate change in facial expression. If the wrong facial expression appeared, they often used strongly exaggerated tone of voice to correct the "misunderstanding." The subjects readily discerned intensity differences in Kismet's expression (reflecting different intensities in the underlying `emotional` state) and modulated their tone of voice to influence them. For instance, small smiles versus large grins were often used to discern how "happy" the robot was. Small ear perks versus widened eyes with elevated ears and craning the neck forward were often used to discern growing levels of "interest" and "attention."

During course of the interaction, several interesting dynamic social phenomena arose. For instance, several of the subjects reported experiencing a very strong emotional response immediately after "successfully" scolding Kismet. In these cases, the robot's saddened face and body posture was enough to arouse a strong sense of empathy. The subject would often immediately stop and look to the experimenter with an anguished expression on her face, claiming to feel "terrible" or "guilty." In this emotional feedback cycle, the robot's own affective response to the subject's vocalizations evoked a strong and similar emotional response in the subject as well. Another interesting social dynamic observed involved *affective mirroring* between robot and human. In this situation, the subject might first issue a medium-strength prohibition to the robot, which causes it to dip its head. The subject responds by lowering her own head and reiterating the prohibition, this time a bit more foreboding. This causes the robot to dip its head even further and look more dejected. The cycle continues to increase in intensity until it bottoms out with both subject and robot having dramatic body postures and facial expressions that mirror the other. This technique was employed to modulate the degree to which the strength of the message was "communicated" to the robot.

6 Summary

We have presented a biologically inspired framework for emotive communication and interaction between expressive anthropomorphic robots and humans. This paper primarily pursues an engineering goal to build a robot that can interact with people in familiar social terms, focusing on affective interactions. However a scientific exploration of the emotion models implemented on Kismet is an interesting possibility for future work. By modeling Kismet's emotional responses after those of living systems, people have a natural and intuitive understanding of Kismet's emotional behavior and how to influence it. From our studies, we have found this to be mutually beneficial for both human and robot. It is beneficial for the robot because it can now socially tune the human's behavior to be appropriate for itself – getting the person to bring the desired stimulus into contact at the appropriate time and at an appropriate intensity. It benefits the human because the person do not require any special training to have a comprehensible and rewarding interaction with the robot – knowing when the robot has understood one's affective state and knowing how one's behavior is influencing the robot's affective state. In general, we have found that expressive feedback plays an important role in facilitating natural and intuitive human-robot communication.

References

1. K. Dautenhahn. Getting to know each other – artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems*, 16(2–4):333–356, December 1995.
2. Juan Velasquez. Modeling emotions and other motivations in synthetic agents. In *Proceedings of the 1997 National Conference on Artificial Intelligence (AAAI97)*, pages 10–15, Rrovidence, RI, 1997.
3. S.Y. Yoon, B. Blumberg, and G. Schneider. Motivation driven learning for interactive synthetic characters. In *Proceedings of the Fourth International Conference on Autonomous Agents (Agents00)*, Barcelona, Spain, 2000.
4. R. Plutchik. *The Emotions.* University Press of America, Lanham, MD, 1991.
5. P. Ekman and H. Oster. Review of research, 1970 to 1980. In P. Ekman, editor, *Emotion in the Human Face*, pages 147–174. Cambridge University Press, Cambridge, UK, 1982.
6. C. Breazeal. *Sociable Machines: Expressive Social Exchange Between Humans and Robots.* PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA, May 2000.
7. C. Smith and H. Scott. A componential approach to the meaning of facial expressions. In J. Russell and J. Fernandez-Dols, editors, *The Psychology of Facial Expression*, pages 229–254. Cambridge University Press, Cambridge, UK, 1997.