Emotion and Sociable Humanoid Robots

Cynthia Breazeal

[a]*MIT Media Lab*
Massachusetts Institute of Technology 77 Massachusetts Ave, NE18-5FL
Cambridge, MA 02139
(617) 452-5601 (ph), (617) 225-2009 (fax)
cynthiab@media.mit.edu

Abstract

This paper focuses on the role of emotion and expressive behavior in regulating social interaction between humans and expressive anthropomorphic robots, either in communicative or teaching scenarios. We present the scientific basis underlying our humanoid robot's emotion models and expressive behavior, and then show how these scientific viewpoints have been adapted to the current implementation. Our robot is also able to recognize affective intent through tone of voice, the implementation of which is inspired by the scientific findings of the developmental psycholinguistics community. We first evaluate the robot's expressive displays in isolation. Next, we evaluate the robot's overall emotive behavior (i.e., the coordination of the affective recognition system, the emotion and motivation systems, and the expression system) as it socially engages nave human subjects face-to-face.

Key words: Human-robot interaction, emotion, expression, sociable humanoid robots.

1 Introduction

Sociable humanoid robots pose a dramatic and intriguing shift in the way one thinks about control of autonomous robots. Traditionally, autonomous robots are designed to operate as independently and remotely as possible from humans, often performing tasks in hazardous and hostile environments (such as sweeping minefields, inspecting oil wells, or exploring other planets). Other applications such as delivering hospital meals, mowing lawns, or vacuuming floors bring autonomous robots into environments shared with people, but human-robot interaction in these tasks is still minimal.

However, a new range of application domains (domestic, entertainment, health care, etc.) are driving the development of robots that can interact and cooperate with people as a partner, rather than as a tool. In the field of human computer interaction (HCI), research by Reeves & Nass (1996) has shown that humans (whether computer experts, lay-people, or computer critics) generally treat computers as they might treat other people. From their numerous studies, they argue that a social interface may be a truly universal interface (Reeves & Nass, 1996). Humanoid robots (and animated software agents) are arguably well suited to this. Sharing a similar morphology, they can communicate in a manner that supports the natural communication modalities of humans. Examples include facial expression, body posture, gesture, gaze direction, and voice. It is not surprising that studies such as these have strongly influenced work in designing technologies that communicate with and cooperate with people as collaborators.

The paper is organized as follows: first, we review a number of related engineering efforts in building computer animated and robotic systems that interact with people in a social manner. Next, we introduce our expressive humanoid robot, Kismet, and highlight our own efforts in building sociable humanoid robots that engage people through expressive social cues (including emotive responses). Section 4 presents those key principles from the theory of emotion and its expression that have inspired the design of our robot's emotion and expression systems. Sections 5 and 6 focus on the computational models of emotion and motivation that have been implemented on Kismet. The next section presents how Kismet's expressive responses are generated algorithmically. Section 8 evaluates the readability of Kismet's facial expressions, and section 9 evaluates the robot's ability to engage people socially through its expressive responses. We conclude the paper with a discussion and summary of results.

2 Embodied Systems That Interact with Humans

There are a number of systems from different fields of research that are designed to interact with people. Many of these systems target different application domains such as computer interfaces, Web agents, synthetic characters for entertainment, or robots for physical labor. In general, these systems can be either embodied (the human interacts with a robot or an animated avatar) or disembodied (the human interacts through speech or text entered at a keyboard). The embodied systems have the advantage of sending paralinguistic communication signals to a person, such as gesture, facial expression, intonation, gaze direction, or body posture. These embodied and expressive cues can be used to complement or enhance the agent's message. At times, para-linguistic cues carry the message on their own, such as emotive facial expressions or gestures. These embodied systems must also address the issue of sensing the human, often focusing on perceiving the human's embodied social cues. Hence, the perceptual problem for these systems is more challenging than that of disembodied systems. This section highlights a few embodied systems, both animated and robotic.

2.1 Embodied Conversation Agents

There are a number of graphics-based systems that combine natural language with an embodied avatar. The focus is on natural, conversational discourse accompanied by gesture, facial expression, and so forth (Cassell, 1999). In some applications, the human uses these systems to perform a task. One of the most advanced systems in this respect is *Rea* from the Media Lab at MIT (Cassell et al., 2000). Rea is a synthetic real-estate agent, situated in a virtual world, that people can query about buying property. The system communicates through speech, intonation, gaze direction, gesture, and facial expression. It senses the location of people in the room and recognizes a few simple gestures. Another significant application area is tutoring systems where the agent helps a person learn how to perform a task. An advanced pedagogical system is *Steve*, developed at USC (Rickel & Johnson, 2000). The human is immersed in virtual reality to interact with the avatar. It supports domain-independent capabilities to support task-oriented dialogs in 3D virtual worlds. For instance, Steve trains people how to operate a variety of equipment on a virtual ship and guides them through the ship to show them where the equipment is located. Sometimes, the task could simply be to communicate with others in a virtual space, a sort of animated "chatroom" with embodied avatars (Vilhjalmsson & Cassell, 1998). There are a number of graphical systems where the avatar predominantly consists of a face with minimal to no body. In Takeuchi & Nagao (1993), for instance, the use of an expressive graphical face to accompany dialogue is explored. They found that the facial component was good for initiating new users to the system, but its benefit was not as pronounced over time. Also of note in this issue is the paper on *Greta* by de Rosis *et al.* that describes how Greta's verbal and non-verbal signals are synchronized and animated in 3D.

2.2 Human-Friendly Robots

The ability to interact with people in the human environment has been a recent motivator of the humanoid robotics community and the service robotics community. For systems such as these, safety and minimizing impact on human living spaces are important issues, as well as the issues of performance and ease of use. For example, the *MOVAID* system (Dario & Susani, 1996) and a similar project at Vanderbilt University (Kawamura et al., 1996) focus on providing assistance to the elderly or to the disabled. In an more educational setting, a number of mobile museum tour guide robots are employed around the world such as *Sage* from the University of Pittsburgh (Nourbakhsh et al., 1999). In the entertainment market, there are a growing number of synthetic pets, one of the best known being Sony's robot dog *Aibo*. Much of the research in the humanoid robotics community has focused on traditional challenges of robot locomotion (e.g., Honda's *P3* bipedal walker (Hirai, 1998)) and upper-torso control for object manipulation tasks (e.g., ATR's humanoid, *DB*). A few humanoid projects have explored the social dimension, such as *Cog* at the MIT AI Lab (Brooks et al., 1999).

2.3 Expressive Face Robots

There are several projects that focus on the development of expressive robot faces, ranging in appearance from being graphically animated (Bruce et al., 2001), to resembling a mechanical cartoon (Takanobu et al., 1999; Scheef et al., 2000), to pursuing a more organic appearance (Hara, 1998; Hara & Kobayashi, 1996). For instance, researchers at the Science University of Tokyo have developed the most human-like robotic faces (typically resembling a Japanese woman) that incorporate hair, teeth, silicone skin, and a large number of control points (Hara, 1998) that map to the *facial action units* of the human face (Ekman & Friesen, 1982). Using a camera mounted in the left eyeball, the robot can recognize and produce a predefined set of emotive facial expressions (corresponding to anger, fear, disgust, happiness, sorrow, and surprise). A number of simpler expressive faces have been developed at Waseda University, one of which can adjust its amount of eye-opening and neck posture in response to light intensity (Takanobu et al., 1999). The robot, *Feelix*, by Canamero & Fredslund (2001) is a Lego-based face robot used to explore tactile and affective interactions with people. It is increasingly common to integrate expressive faces with mobile robots that engage people in an educational or entertainment setting, such as museum tour guide robots (Nourbakhsh et al., 1999; Burgard et al., 1998).

As expressive faces are incorporated into service or entertainment robots, there is a growing interest in understanding how humans react to and interact with them. For instance, Kiesler & Goetz (2002) explored techniques for characterizing people's mental models of robots and how this is influenced by varying the robot's appearance and dialog to make it appear either more playful and extraverted or more caring and serious. Bruce et al. (2001) investigated people's willingness to engage a robot in a short interaction (i.e., taking a poll) based on the presence or absence of an expressive face and the ability to indicate attention.

3 Kismet and the Sociable Machines Project

The ability for people to naturally communicate with such machines is important. However, for suitably complex environments and tasks, the ability for people to intuitively teach these robots will also be important. Ideally the robot could engage in various forms of social learning (imitation, emulation, tutelage, etc.), so that one could teach the robot just as one would teach another person. Learning by demonstration to acquire physical skills such as pole balancing (Atkeson & Schaal, 1997*a*,*b*; Schaal, 1997), learning by imitation to acquire a proto-language (Billard, 2002), and learning to imitate in order to produce a sequence of gestures (Demiris & Hayes, 2002; Mataric, 2000) have been explored on physical humanoid robots and physics-based animated humanoids. Although current work in imitation-based learning with humanoid robots has dominantly focused on articulated motor coordination, social and emotional aspects can play a profound role in building robots that can communicate with and learn from people.

The *Sociable Machines Project* develops an expressive anthropomorphic robot called Kismet (see Figure 1) that engages people in natural and expressive face-to-face interaction. An overview of the project can be found in Breazeal (2002*a*). The robot is about 1.5 times the size of an adult human head and has a total of 21 degrees of freedom (DoF). Three degrees of freedom direct the robot's gaze, another three control the orientation of its head, and the remaining fifteen move its facial features (e.g., eyelids, eyebrows, lips, and ears). To visually perceive the person who interacts with it, Kismet is equipped with a total of four color CCD cameras (there is one narrow field of view camera behind each pupil and the remaining two wide field of view cameras are mounted between the robot's eyes as shown). In addition, Kismet has two small microphones (one mounted on each ear). A lavalier microphone worn by the person is used to process their vocalizations.

Inspired by infant social development, psychology, ethology, and evolution, this work integrates theories and concepts from these diverse viewpoints to enable Kismet to enter into natural and intuitive social interaction with a human

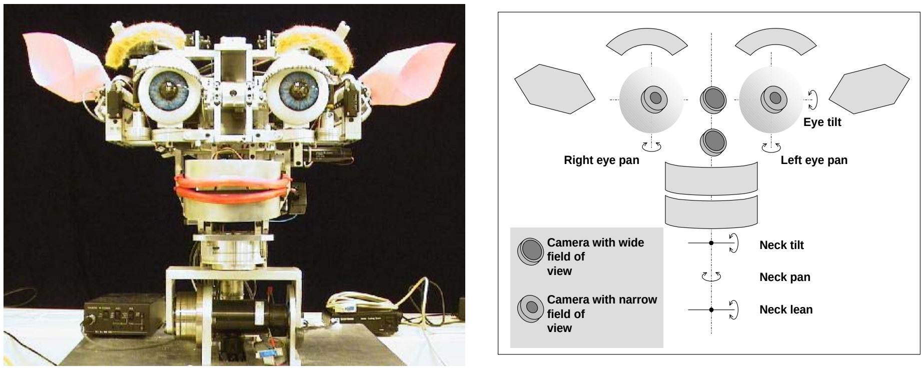

Fig. 1. Kismet, our sociable robot.

and to eventually learn from them, reminiscent of parent-infant exchanges. To do this, Kismet perceives a variety of natural social cues from visual and auditory channels, and delivers social signals to the human through gaze direction, facial expression, body posture, and vocal babbles. The robot has been designed to support several social cues and skills that could ultimately play an important role in socially situated learning with a human instructor. These capabilities are evaluated with respect to the ability of naive subjects to read and interpret the robot's social cues, the robot's ability to perceive and appropriately respond to human social cues, the human's willingness to provide scaffolding to facilitate the robot's learning, and how this produces a rich, flexible, dynamic interaction that is physical, affective, social, and affords a rich opportunity for learning.

This paper focuses on the role of emotion and expressive behavior in social interaction between robots and humans. We present Kismet's computational models of emotion, how it expresses its "emotive" state, and how the robot uses these systems in conjunction with its perceptual and behavioral systems to regulate its interaction with people (Breazeal, 1998; Breazeal & Aryananda, 2002). The robot's affective systems are evaluated through studies with naive subjects (Breazeal & Aryananda, 2002; Breazeal, 2002*b*).

4 Emotions and their Expression in Living Systems

Emotions are an important motivation system 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 (Plutchik, 1991; Izard, 1977). For instance, Frijda (1994*a*) suggests that positive emotions are elicited by events that satisfy some motive, enhance one's power of survival, or demonstrate the successful exercise of one's capabilities. Positive emotions often signal that activity toward the goal can terminate, or that resources can be freed for other exploits.

In contrast, many negative emotions result from painful sensations or threatening situations. Negative emotions motivate actions to set things right or to prevent unpleasant things from occurring.

4.1 Theory of Basic Emotions

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 (Ekman, 1992; Izard, 1993). The emotions of anger, disgust, fear, joy, sorrow, and surprise are often supported as being basic from evolutionary, developmental, and cross-cultural studies (Ekman & Oster, 1982). Each basic emotion is posited to serve a particular function (often biological or social), arising in particular contexts, to prepare and motivate a creature to respond in adaptive ways. They serve as important reinforcers for learning new behavior. In addition, emotions are refined and new emotions are acquired throughout emotional development. Social experience is believed to play an important role in this process (Ekman & Oster, 1982).

Several theorists argue that emotion has evolved as a relevance-detection and response-preparation system. They posit an appraisal system that assesses the perceived antecedent conditions with respect to the organism's well-being, its plans, and its goals (Levenson, 1994; Izard, 1994; Frijda, 1994*c*; Lazarus, 1994). Scherer (1994) has studied this assessment process in humans and suggests that people affectively appraise events with respect to novelty, intrinsic pleasantness, goal/need significance, coping, and norm/self compatibility . Hence, the level of cognition required for appraisals can vary widely.

These appraisals (along with other factors such as pain, hormone levels, drives, etc.) evoke a particular emotion that recruits response tendencies within multiple systems. These include physiological changes (such as modulating arousal level via the autonomic nervous system), adjustments in subjective experience, elicitation of behavioral response (such as approach, attack, escape, etc.), and displaying expression. The orchestration of these systems represents a generalized solution for coping with the demands of the original antecedent conditions. Plutchik (1991) calls this stabilizing feedback process *behavioral homeostasis*. Through this process, emotions establish a desired relation between the organism and the environment that pulls the creature toward certain stimuli and events and pushes it away from others. Much of the relational activity can be social in nature, motivating proximity seeking, social avoidance, chasing off offenders, etc. (Frijda, 1994*b*).

The expressive characteristics of emotion in voice, face, gesture, and posture serve an important function in communicating emotional state to others. Levenson (1994) argues that 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. Darwin (1872) argued that 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. As argued by Scherer (1994), 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.

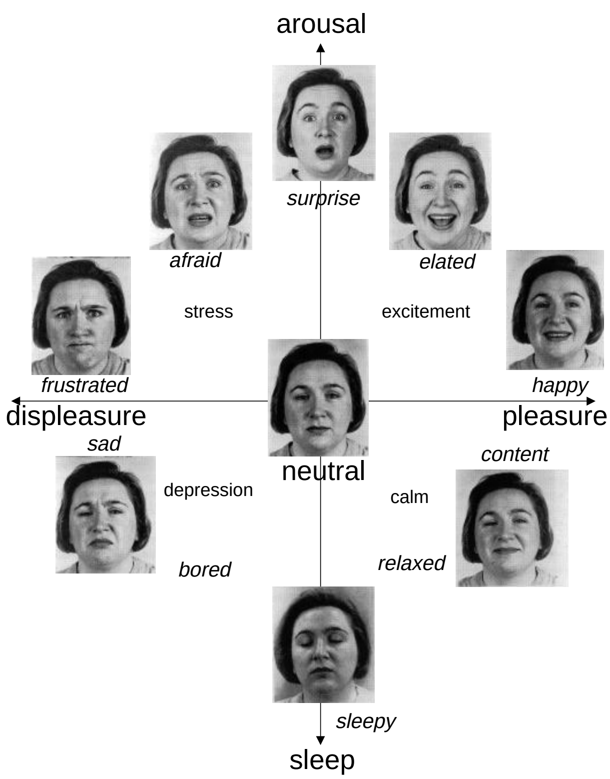

Fig. 2. Russell's pleasure-arousal space for facial expression. Adapted from Russell (1997).

	Facial Action							
Meaning	Eyebrow Frown	Raise Eyebrows	Raise upper Eyelid	Raise Lower Eyelid	Up Turn Lip Corners	Open Mouth	Tighten Mouth	Raise Chin
Pleasantness	⬇				⬆	⬆	⇩	⬇
Goal Obstacle/Discrepancy	⬆							
Anticipated Effort	⇧							
Attentional Activity		⬆	⬆					
Certainty		⇩		⇧		⇧		
Novelty		⇧	⇧					
Personal Agency/Control		⇩	⇩			⇩		

Table 1
A possible mapping of facial movements to affective dimensions proposed by Smith & Scott (1997). An up arrow indicates that the facial action is hypothesized to increase with increasing levels of the affective meaning dimension. A down arrow indicates that the facial action increases as the affective meaning dimension decreases. For instance, the lip corners turn upwards as "pleasantness" increases, and lower with increasing "unpleasantness."

4.2 Componential Approaches to Facial Expression

Instead of viewing emotions in terms of categories (happiness, anger, fear, etc.), one school of thought is to conceptualize the dimensions that could span the relationship between different emotions (arousal and valence, for instance). Psychologists of this view posit that facial expressions have a systematic, coherent, and meaningful structure that can be mapped to affective dimensions (Russell, 1997; Lazarus, 1991; Plutchik, 1984; Smith, 1989; Woodworth, 1938). See figure 2 for an example.

Instead of taking a production-based approach to facial expression (how do emotions generate facial expressions), Russell (1997) takes a perceptual stance (what information can an observer read from a facial expression). Hence, by considering the individual facial action components that contribute to that structure, it is possible to reveal much about the underlying properties of the emotion being expressed. It follows that some of the individual features of expression have inherent signal value. This promotes a signaling system that is robust, flexible, and resilient (Smith & Scott, 1997). It allows for the mixing of these components to convey a wide range of affective messages, instead of being restricted to a fixed pattern for each emotion. This variation allows fine-tuning of the expression, as features can be emphasized, de-emphasized, added, or omitted as appropriate. Furthermore, it is well-accepted that any emotion can be conveyed equally well by a range of expressions, as long as those expressions share a family resemblance. The resemblance exists because the expressions share common facial action units. The facial action units characterize how each facial muscle (or combination of facial muscles) adjust the skin and facial features to produce human expressions and facial movements (Ekman & Friesen, 1982). It is also known that different expressions for different emotions share some of the same face action components (the raised brows of fear and surprise, for instance). It is hypothesized by Smith & Scott (1997) that those features held in common assign a shared affective meaning to each facial expression. The raised brows, for instance, convey attentional activity for both fear and surprise.

5 Models of Emotion and Drives for a Sociable Robot

Kismet's motivations (i.e., its "drives" and "emotions") establish its nature by defining its "needs" and influencing how and when it acts to satisfy them. As a convention, we use a different font to distinguish parts of the architecture of this particular system from the general uses of this word. For instance, `emotion` refers to the particular set of computational processes that are active in the system. When the word "emotion" appears with quotes, we are using it as an analogy to emotions in animals or humans.

The nature of Kismet is to socially engage people and ultimately to learn from them. Kismet's `emotion` and `drive` processes are designed such that the robot is in an alert and mildly positive valenced state when it is interacting well with people and when the interactions are neither overwhelming nor understimulating. This corresponds to an environment that affords high learning potential as the interactions slightly challenge the robot yet also allow Kismet to perform well.

5.1 Overview of the Drive System

The design of Kismet's "drives" (Breazeal, 1998) is heavily inspired by ethological views of the analogous process in animals (although it is a simplified and idealized model) (McFarland & Bosser, 1993). At any point in time, the robot's behavior is organized about satiating its "drives". Each `drive` is modeled as a separate process, shown in Figure 3. There are three drives implemented on Kismet that establish the top-level goals of the robot: to engage people, to engage toys, and to occasionally rest. One distinguishing feature of a `drive` is its temporally cyclic behavior. That is, given no stimulation, a `drive` will tend to increase in intensity unless it is satiated. This is analogous to an animal's degree of hunger or level of fatigue, both following a cyclical pattern. Another distinguishing feature is its homeostatic nature. Each acts to maintain a level of intensity within a bounded range—neither too much nor too little. Its change in intensity reflects the ongoing "needs" of the robot and the urgency for tending to them.

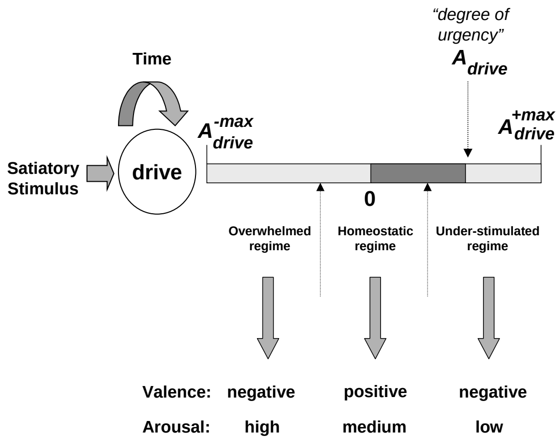

Fig. 3. The homeostatic model of a drive process. The circle in the figure represents the drive as a computational process with inputs (shown by the arrows) and an internal representation of intensity (shown by the thermometer-style bar). Each drive has a temporal input to implement its cyclic behavior as shown in the figure. The activation energy A_{drive} of each `drive` ranges between $[A_{drive}^{-max}, A_{drive}^{+max}]$, where the magnitude of the A_{drive} represents its intensity. For a given A_{drive} intensity, a large positive magnitude corresponds to under-stimulation by the environment, whereas a large negative magnitude corresponds to over-stimulation by the environment. The level of the drive returns to the homeostatic regime when the robot encounters the satiatory stimulus. Each `drive` can influence the robot's "emotional" state according to how well the robot's "needs" are being met (see section 6.2).

There is a desired operational point and acceptable bounds of operation around that point. In general, the activation level of each `drive` is partitioned into three regimes: an *under-stimulated regime*, an *overwhelmed regime*, and a *homeostatic regime*. A drive remains in its homeostatic regime when it is encountering its satiatory stimulus and that stimulus is of appropriate intensity. In the absence of the satiatory stimulus (or if the intensity is too low), the `drive` tends toward the under-stimulated regime. Alternatively, if the satiatory stimulus is too intense (e.g., moving too close or too fast), the `drive` tends toward the overwhelmed regime. Hence, to remain in balance, it is not sufficient that the satiatory stimulus be present; it must also be of an appropriate intensity.

Kismet's `drives` serve several purposes. They influence behavior selection by directly and preferentially passing activation to some behaviors over others (i.e., those that serve to satiate the drive). They also provide a functional context (i.e., the goal, namely which "need" the robot is actively trying to address) that organizes behavior and perception. Furthermore, they influence the robot's affective state by directly contributing to valence and arousal measures as shown in Figure 3 (the details of which are presented in section 6.2).

Thus, the `drives` can indirectly bias behavior through the emotion system as well. Since the `drives` operate on a slower time scale than the `emotions`, they contribute to the the long-term affective state (or "mood" of the robot) and its expression.

In the current implementation there are three `drives`. The `social drive` motivates the robot to be in the presence of people and to interact with them. On the under-stimulated extreme, the robot is "lonely"; it is predisposed to act in ways to establish face-to-face contact with people. On the overwhelmed extreme, the robot is predisposed to act in ways to avoid face-to-face contact (e.g., when a person is over-stimulating the robot by either moving too much or being too close to the robot's eyes). In similar manner, the `stimulation drive` motivates the robot to interact with things, such as colorful toys. The `fatigue drive` is unlike the others in that its purpose is to allow the robot to shut out the external world instead of trying to regulate its interaction with it. While the robot is "awake," it receives repeated stimulation from the environment or from itself. As time passes, this `drive` approaches the "exhausted" end of the spectrum. Once the intensity level exceeds a certain threshold, it is time for the robot to "sleep." While the robot sleeps, *all* `drives` return to their homeostatic regimes, allowing the robot to satiate its drives if the environment offers no significant stimulation.

5.2 Overview of the Emotion System

The organization and operation of the *emotion system* is strongly inspired by various theories of emotions in humans. In concert with the robot's `drives`, it is designed to be a flexible system that mediates between both environmental and internal stimulation to elicit an adaptive behavioral response that serves either social or self-maintenance functions. The `emotions` are triggered by various events that are evaluated as being of significance to the "well-being" of the robot. Once triggered, each `emotion` serves a particular set of functions to establish a desired relation between the robot and its environment. They motivate the robot to come into contact with things that promote its "well-being" and to avoid those that do not. In general, our implementation is strongly inspired by ethological models for perception, motivation, and behavior (Tinbergen, 1951; Lorenz, 1973). Consequently, at a high level our emotion system is similar in spirit to the *Cathexis* system of Velasquez (1996). The significant differences with Cathexis are discussed in Section 10.

As shown in Table 2, a number of emotive responses have been implemented on Kismet. It summarizes under what conditions certain `emotions` and behavioral responses arise, and what function they serve the robot. This table is derived from the evolutionary, cross-species, and social functions hypothesized

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 2
Summary of the antecedents and behavioral responses that comprise Kismet's emotive responses. The antecedents refer to the eliciting perceptual conditions for each `emotion`. The behavior column denotes the observable response that becomes active with the `emotion`. For some this is simply a facial expression. For others, it is a behavior such as `escape`. Note that behaviors as well as `drives` encode the goals of the robot. For instance, the goal of the `escape` behavior is to move away from a threatening stimulus. The `stimulation` drive may bias the activation of this behavior when its level of intensity moves to the overwhelmed regime due to the presense of a close and fast moving object. The column to the right describes the function each emotive response serves Kismet.

by Plutchik (1991), Darwin (1872), and Izard (1977). It includes the six primary emotions proposed by Ekman (1992). There are several processes in the emotion system that model different arousal states (such as interest, calm, or boredom) that also have a corresponding expression and a few have an associated behavioral response. These emotive responses map well to several proto-social responses of human infants, and hence are of particular relevance to Kismet's design (Breazeal & Scassellati, 1999).

There are three systems that contribute to the goals of the robot. The homeostatic goal of each `drive` can be conceptualized as survival-based goals. Kismet therefore has a goal to interact with people, a goal to be stimulated by toys, and to occasionally rest. The degree to which each drive is satiated in a timely fashion contributes to the robot's overall measure of its "well being." The emotion system contributes to the goals of bringing the robot into contact with things that benefit it and to avoid those things that are undesirable or potentially harmful. Each emotive response summarized in Table ?? performs this in its own distinct manner through behavioral homeostasis. Whereas the `drives` address a long-term measure of the "well being" of the robot, the emotive responses work on a faster time scale. The behavior system consists of a hierarchy of task-based goals and is organized in the spirit of those ethological models proposed by Tinbergen (1951) and Lorenz (1973). Each behavior coordinates sensori-motor patters to achieve a particular task such as search behaviors, approach behaviors, avoidance behaviors, and interaction behaviors. Both the `drives` and `emotions` use the behavior system as a resource to carry out their own goals by biasing an appropriate behavioral response to become active at the right time.

By adapting these ideas to Kismet, the robot's emotional responses mirror those of biological systems and therefore should seem plausible to a human. This is very important for social interaction because it makes the robot's emotive responses and their reasons for coming about consistent with what a human might expect. Each of the entries in this table has a corresponding affective display. For instance, the robot exhibits sorrow upon the prolonged absence of a desired stimulus. This may occur if the robot has not been engaged with a toy for a long time. The sorrowful expression is intended to elicit attentive acts from the human. 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 stern 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). In addition, the expression on the robot's face is a social signal to the human, who responds in a way to further promote the robot's "well-being." 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.

6 Components of Emotion

Several theories posit that emotional reactions consist of several distinct but interrelated facets (Scherer, 1984; Izard, 1977). In addition, several appraisal theories hypothesize that a characteristic appraisal (or meaning analysis) triggers the emotional reaction in a context sensitive manner (Frijda, 1994*b*; Lazarus, 1994; Scherer, 1994). 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

In living systems, it is believed that these individual facets are organized in a highly interdependent fashion. Physiological activity is hypothesized to physically prepare the creature to act in ways motivated by action tendencies. Furthermore, both the physiological activities and the action tendencies are organized around the adaptive implications of the appraisals that elicited the emotions. From a functional perspective, Smith (1989) and Russell (1997) suggest that the individual components of emotive facial expressions are also linked to these emotional facets in a highly systematic fashion.

6.1 Emotive Releasers

We begin this discussion with the input to the *emotion system* (see figure 4). External events, such as visual and auditory stimuli, are sensed by the robot and are filtered by a number of feature extractors (e.g., color, motion, pitch, etc.). In the *high level perceptual system*, these features are bound by *releaser* processes that encode the robot's current set of beliefs about the state of the robot and its relation to the world.

There are many different kinds of releasers defined for Kismet, each hand-crafted, and each combining different contributions from a variety of factors. The activation level of a given releaser process rises above threshold when all of its perceptual and internal conditions are present with sufficient intensity. Each releaser can be thought of as a simple "cognitive" assessment that combines lower-level perceptual features with measures of its internal state into behaviorally significant perceptual categories. These include attributes such as the presence or absence of a stimulus (and for how long), its nature (e.g., toy-related or person-related), the quality of the stimulus (e.g., the intensity is too low, too high, or just right), or whether it is desired or not (e.g., it relates to the active goals or motivations). For instance, the features of `color`, `size`, `motion`, `proximity` are integrated to form a toy percept. If the `stimulation drive` is being tended to and the toy is neither too fast nor too close to the robot, then the `desired-toy` releaser is active. However, if the `social drive` is being tended to instead, then the `undesired-toy` releaser is active. If the toy has an aggressive motion (i.e., too close and moving too fast), then the `threatening-toy` releaser is. Many of these are antecedent conditions that are tailored to specific emotive responses.

Hence, each releaser is evaluated with respect to the robot's "well-being" and

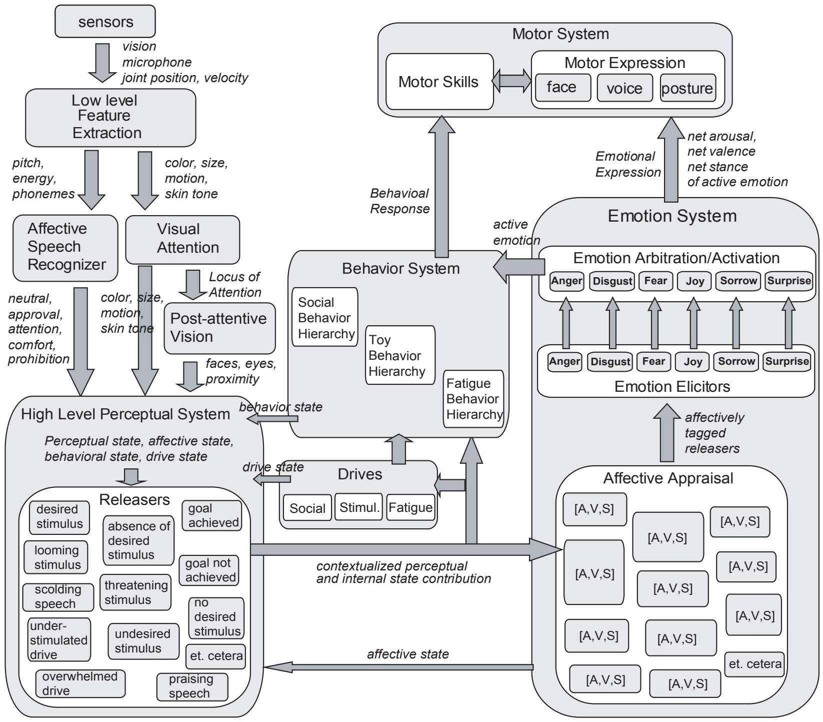

Fig. 4. An overview Kismet's cognitive architecture. External events, such as visual and auditory stimuli, are sensed by the robot and are filtered by a number of feature extractors (e.g., color, motion, pitch, etc.). In the high level perceptual system, these features are bound by releaser processes that encode the robot's current set of beliefs about the internal and external state of the robot and its relation to the world. The result is a set of response-specific releasers that serve as antecedent conditions for specific emotive responses. The active releasers are passed to an affective appraisal phase where they are tagged with affective information (i.e., arousal, valence, and stance as denoted by $[A, V, S]$ in the figure). The tagging process is discussed in section 6.2 . All active contributions from the affective assessment phase are filtered through the *emotion elicitors* for each emotion process. In the *emotion arbitration* phase, the `emotion` processes compete for activation in a winner-take-all scheme. The winner evokes a corresponding facial expression, body posture, and vocal quality by sending the net $[A, V, S]$ to the expressive motor system. The winner may also evoke a corresponding behavioral response by sending activation energy to the corresponding behavior in the behavior system (such as `flee` in the case of `fear` as shown in Figure 6).

its goals. This evaluation is converted into an activation level for that releaser. If the perceptual features and evaluation are such that the activation level is above threshold (i.e., the conditions specified by that releaser hold), then its output is passed to its corresponding behavior process in the behavior system. It is also passed to the affective appraisal stage where it can influence the emotion system. There are a number of factors that contribute to the assessment made by each releaser. They are as follows:

- *Drives*. The active `drive` provides important context for many releasers.

In general, it determines whether a given type of stimulus is either desired or undesired. For instance, if the `social drive` is active, then skin-toned stimuli are desirable, but colorful stimuli are undesirable (even if they are of good quality). Hence, this motivational context plays an important role in determining whether the emotional response will be one of incorporation or rejection of a presented stimulus. In addition, there is a releaser defined for each regime of each `drive` to represent how well each `drive` is being satiated.

- *Affective State.* The current affective state also provides important context for certain releasers. A good example is the `soothing-speech` releaser. Given a "soothing" classification from the affective speech recognizer (Breazeal & Aryananda, 2002), the `soothing-speech` releaser only becomes active if Kismet is already distressed. Otherwise, the `neutral-speech` releaser is activated. This second stage of processing reduces the number of misclassifications between "soothing" speech versus "neutral" speech.
- *Active Behavior(s).* The behavioral state plays an important role in disambiguating certain perceptual conditions as well. For instance, a `no-face` perceptual condition could correspond to several different possibilities. The robot could be engaged in a `seek-people` behavior, in which case a skin-toned stimulus is a desired but absent stimulus. Initially this could be encoded in a `missing-desired-face` releaser which would trigger exploration behavior. Over time, however, this could contribute to another releaser, the `prolonged-absence-desired-face` releaser, that signals a state of deprivation due to a long-term loss. Alternatively, the robot could be engaged in an `escape` behavior. In this case, `no-face` corresponds to successful escape (signaled by the `threatening-face-gone` releaser), a rewarding circumstance. In addition, there are a set of releasers defined for each behavior that indicate whether its goal has been achieved or not, and if not then for how long.

6.2 Affective Appraisal

Each releaser with activation above threshold is appraised in affective terms by an associated *somatic marker* (SM) process in the *affective appraisal* stage. This mechanism is inspired by Damasio's *Somatic Marker Hypothesis* where incoming perceptual, behavioral, or motivational information is "tagged" with affective information (Damasio, 1994). There are three classes of tags the SM uses to affectively characterize a given releaser. Each tag has an associated value (ranging from -1250 to +1250)that represents 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 releaser 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 to the robot. Positive values correspond to advance whereas negative values correspond to retreat. Hence a releaser that corresponds to a threatening stimulus, such as the `threatening-toy` releaser, would be assigned affective tags with values of $A = 1200$ (very arousing), $V = -1000$ (very unfavorable), and $S = -1000$ (strong avoidance).

There are three systems within Kismet's cognitive architecture (other than the emotion system) that contribute to its net affective state by way of their associated releaser(s) (note the arrows from these systems to the releasers in Figure 4) and the somatic marking processes. First, there are external environmental factors that come by way of the high-level perceptual system that elicit emotive responses (such as praising speech or a threatening stimulus). Next, within the motivation system, each regime of each drive biases arousal and valence differently. This in turn contributes to the activation of different `emotion` processes. The homeostatic regime is marked with positive valence and balanced arousal, contributing to a "contented" affective state. The understimulated regime is marked with negative valence and low arousal, contributing to a "bored" affective state that can eventually decline to "sorrow." The overwhelmed regime is marked with negative valence and high arousal, contributing to an affective state of "distress." Hence, the robot's affective state becomes less desirable when its "need" to interact with people and to be stimulated with toys are not adequately met. Finally, within the behavior system, the success or delayed progress of the active behavior toward its goal (encoded by the `success-achieved` and `level-of-frustration` releasers) can also influence the affective state. Success in achieving the current goal is marked with positive valence, whereas delayed progress is marked with negative valence.

In general, there are four factors that the designer considers when assigning respective values to $[A, V, S]$:

- *Intensity.* One important factor is how intense a stimulus is to the robot. Stimuli that are closer to the robot, move faster, or are larger in the field of view are more intense than stimuli that are further, slower, or smaller. The intensity of the stimulus generally maps to arousal. Releasers that represent threatening or very intense stimuli are assigned a high arousal value. Absent or low intensity stimuli are assigned a low arousal value. Soothing speech has a calming influence on the robot, so it also serves to lower arousal if initially high.
- *Relevance.* The relevance of the stimulus (whether it addresses the current goals of the robot) influences the values assigned to valence and stance. Stimuli that are relevant are "desirable" and are assigned a positive valence value and an approaching stance value. Stimuli that are not relevant

are “undesirable” and are assigned with negative arousal and withdrawing stance values.

- *Intrinsic Affect of Stimuli.* Some stimuli are hardwired to influence the robot’s affective state via the releasers in a specific manner. Praising speech is assigned positive valence and slightly high arousal values. Scolding speech is assigned with negative valence and low arousal values(tending to elicit `sorrow`). Attentional bids alert the robot and are assigned with a medium arousal value. Looming stimuli startle the robot and are assigned a high arousal value. Threatening stimuli are assigned with high arousal, negative valence, and withdrawing stance, thereby contributing to a fear response.
- *Goal Directedness.* Each behavior in the behavior system specifies a task-achieving goal, i.e., a particular relation the robot wants to maintain with the environment. Given an active behavior, there are two special releasers that reflect the active behavior’s internal measure of progress towards its goal. Success in achieving a goal (as represented by the `success-achieved` releaser) is assigned with a value of positive valence and thereby promotes `joy`. Prolonged delay in achieving a goal (denoted by the `level-of-frustration` releaser) is assigned with values of negative valence and withdrawn stance. The value of the stance component increases slowly over time (from withdrawn to approaching) to encourage an emotive transition to `anger` the longer the robot fails to achieve its goal.

Because there are potentially many different kinds of factors that modulate the robot’s affective state (e.g., behaviors, motivations, perceptions), 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. In the current implementation, the affective tags for each releaser are specified by the designer. These may be fixed constants, or linearly varying quantities.

6.3 Emotion Elicitors

All somatically marked inputs are passed to the *emotion elicitor* stage. Each affect process, that is, each emotion type, has an associated $[A, V, S]$ profile (see Figure 5 which summarizes how $[A, V, S]$ values map onto each `emotion` process.). The purpose of the emotion elicitor stage is to determine which of the tagged releasers contribute to the activation of the distinct affective processes (e.g., `anger`, `fear`, etc.). 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 `sorrow` 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.

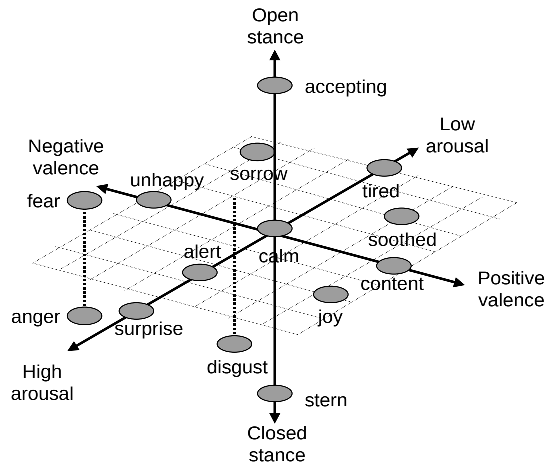

Fig. 5. Mapping of emotional categories to arousal, valence, and stance dimensions $[A, V, S]$.

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 5. 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.

Several different schemes for computing the net contribution to a given `emotion` process were tried. The scheme described above was selected because its properties yielded the most appropriate behavior. In an earlier version, all the incoming contributions were simply averaged. This tended to "smooth" the net affective state to an unacceptable degree. For instance, if the robot's `fatigue-drive` was high (biasing a low arousal state) and a threatening toy appeared (contributing to a strong negative valence and high arousal), the averaging technique resulted in a slightly negative valence and neutral arousal. This is insufficient to evoke `fear` and an escape response when the robot should have protected itself.

As an alternative, we could have hard-wired certain releasers directly to `emotion` processes. It is not clear, however, how this approach supports the influence of `drives` and behaviors, whose affective contributions change as a function of time. For instance, a given `drive` contributes to `fear`, `sorrow`, or `interest` processes depending on its current activation regime. The current approach balances the constraints of having certain releasers contribute heavily and directly to the appropriate emotive response, while accommodating those influences that contribute to different `emotions` as a function of time. The end result also has important implications for generating facial expressions that reflect this assessment process in a rich way. This is important for social interaction as originally argued by Darwin (1872). This expressive benefit is discussed in further detail in Section 7.

6.4 Emotion Activation

Next, the activation level of each `emotion` process is computed. A separate process exists for each `emotion` defined in Table 2; That is, for `joy`, `anger`, `disgust`, `fear`, `sorrow`, `surprise`, `interest`, `boredom`, and `calm`.

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 (also determined empirically) before they are expressed externally. The activation of each process is computed by the equation:

$$A_{emotion} = E_{emotion} + B_{emotion} + P^{active}_{emotion} - \delta_t$$

where $E_{emotion}$ is the activation level of its affiliated elicitor process, $B_{emotion}$ is a constant offset particular to this `emotion` that can be used to make it easier to activate. $P^{active}_{emotion}$ is another bias term that adds a level of persistence to the `emotion` when it becomes active. In essence, it 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, unlike `drives` (which contribute to the robot's longer-term "mood"), the `emotions` have an intense expression followed by decay to a baseline intensity. The decay takes place on the order of seconds. The $B_{emotion}$, $P^{active}_{emotion}$, and δ_t values and decay rates are determined empirically. Hence, whereas the $E_{emotion}$ computes the relevance of a particular `emotion`, the other terms determine its temporal qualities (how easy it is to activate and for how long it remains active).

6.5 Emotion Arbitration

Next, the `emotion` processes compete for control in a winner-take-all arbitration scheme based on their activation level. The activation level of an `emotion` process is a measure of its relevance to the current situation. Each of these processes is distinct from the others and regulates the robot's interaction with its environment in a distinct manner. Each becomes active in a different environmental (or internal) situation. Each motivates a different observable response by spreading activation to a specific behavior process in the behavior system. If this amount of activation is strong enough, then the active emotion can "seize" temporary control and force the behavior to become expressed. In a process of behavioral homeostasis (Plutchik, 1991), the emotive response maintains activity through feedback until the correct relation of robot to environment is established.

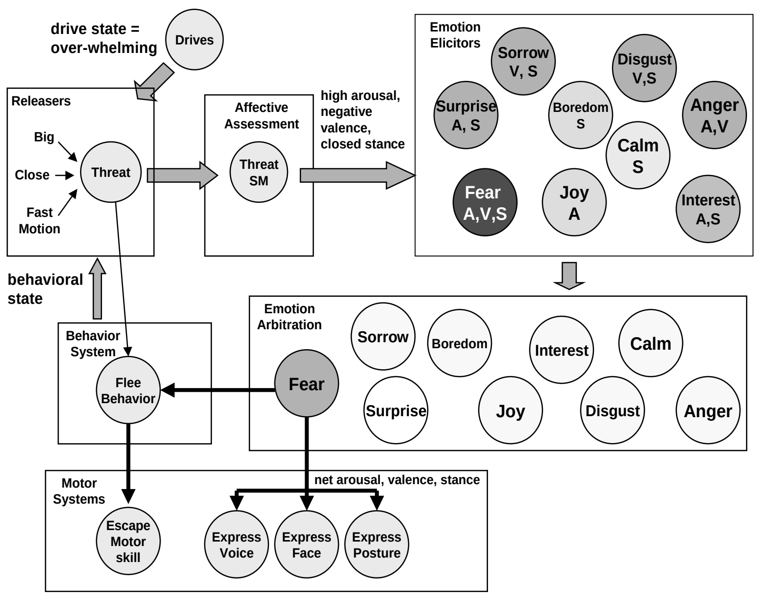

Fig. 6. The implementation of the `fear` emotive response. The releaser for `threat` is passed to the affective appraisal phase. It is tagged with high arousal, negative valence, and closed stance by the corresponding somatic marker process in the affective assessment stage. This affective information is then filtered by the corresponding elicitor of each emotion process. Darker shading corresponds to a higher activation level. Note that only the `fear elicitor` process has each of the arousal, valence, and stance conditions matched (hence, it has the darkest shading). As a result, it is the only one that passes activation to its corresponding `emotion` process. When `fear` is active, it sends activation energy to evoke the `flee` behavior. The goal of the `flee` behavior is to protect the robot from a damaging stimulus. To achieve this goal, it first displays an expressive component where the net $[A, V, S]$ is sent to the expressive motor system. This generates the corresponding facial expression, body posture, and vocal quality. If this communicative signal does not stop the threatening stimulus, the `escape` motor skill is activated, which causes the robot to close its eyes and turn its head away from the threatening stimulus.

Concurrently, the net $[A, V, S]$ of the active process is sent to the expressive components of the motor system, causing a distinct facial expression, vocal quality, and body posture to be exhibited. The strength of the facial expression reflects the level of activation of the `emotion`. Figure 6 illustrates the emotional response network for the `fear` emotion process. Affective networks for the other responses in Table 2 are defined in a similar manner. People have a natural and intuitive understanding of Kismet's emotional behavior since it adheres to their expectations for those of living creatures.

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 person 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 person gets immediate expressive feedback that she is "frightening" the robot. If this was not the intent, then the person has an intuitive understanding of why the robot is frightened and modifies his/her behavior accordingly. The facial expression also sets the human's expectation of what behavior will soon follow. As a result, the human not only sees what the robot is doing, but has an understanding of why.

7 Generating Emotive Expression

The emotion system influences the robot's facial expression. The human can read the robot's facial expression to interpret whether the robot is "distressed" or "content," and can adjust his/her interactions with the robot accordingly. The person accomplishes this by adjusting either the type (social versus non-social) and/or the quality (low intensity, moderate intensity, or high intensity) of the stimulus presented to Kismet. These emotive cues are critical for helping the human work with the robot to establish and maintain a suitable interaction where the robot's drives are satisfied, where it is sufficiently challenged, yet where it is largely competent in the exchange.

As discussed in section 4, Russell (1997) argues the human observer perceives two broad affective categories on the face, arousal and pleasantness. As shown in Figure 2, he maps several emotions and corresponding expressions to these two dimensions. This scheme, however, seems fairly limiting for Kismet. First, it is not clear how all the primary emotions are represented with this scheme (disgust is not accounted for). It also does not account for positively valenced yet reserved expressions such as a coy smile or a sly grin (which hint at a behavioral bias to withdraw). More importantly, "anger" and "fear" reside in very close proximity to each other despite their very different behavioral correlates. From an evolutionary perspective, the behavioral correlate of anger is to attack (a very strong approaching behavior), and the behavioral correlate for fear is to escape (a very strong withdrawing behavior). These are stereotypical responses derived from cross-species studies — obviously human behavior can vary widely. Nonetheless, from a practical engineering perspective of *generating* expression, it is better to separate these two emotional responses by a greater distance to minimize accidental activation of one instead of the other. Adding the stance dimension addressed these issues for Kismet.

7.1 Affect Space

Kismet's facial expressions are generated using an interpolation-based technique over a three-dimensional space. The three dimensions correspond to arousal, valence, and stance. Recall in Section 5, the same three attributes are used to affectively assess the myriad of environmental and internal factors that contribute to Kismet's affective state. We call the space defined by the $[A, V, S]$ trio the *affect space*. The current affective state occupies a single point in this space at a time. As the robot's affective state changes, this point moves about within this space. Note that this space not only maps to emotional states (e.g., anger, fear, sorrow, etc.) but also to the level of arousal as well (e.g., excitement and fatigue). The affect space can be roughly partitioned into regions that map to each emotion process (see Figure 5).

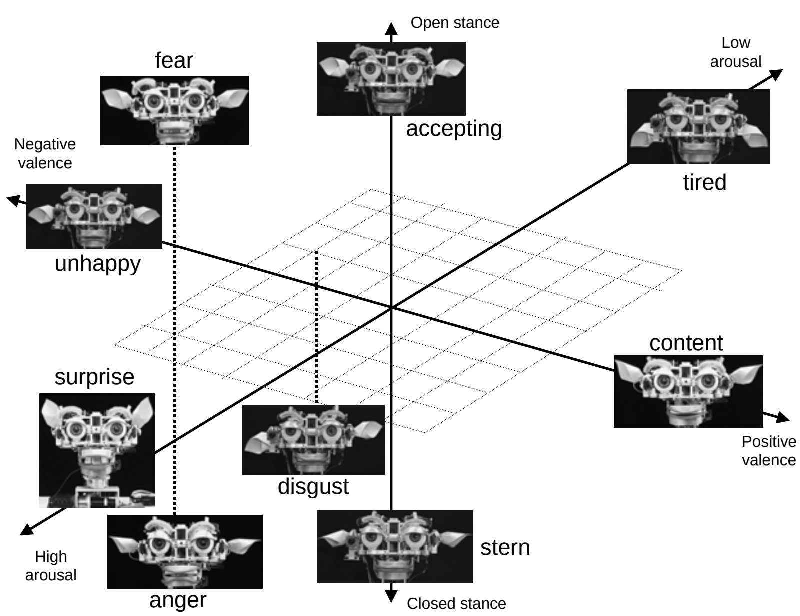

Fig. 7. This diagram illustrates where the basis postures are located in affect space.

7.2 *Basis Postures*

There are nine *basis* (or *prototype*) postures that collectively span this space of emotive expressions (see Figure 7). Although some of these postures adjust specific facial features more strongly than the others, each prototype influences most if not all of the facial features to some degree. For instance, the valence prototypes have the strongest influence on lip curvature, but can also adjust the positions of the ears, eyelids, eyebrows, and jaw. 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 in order 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, surprise, and more. Different levels of arousal can be expressed as well from interest, to calm, to weariness.

The primary six prototype postures sit at the extremes of each dimension (see Figure 7). They correspond to high arousal, low arousal, negative valence, positive valence, open (approaching) stance, and closed (withdrawing) stance. The high arousal prototype, P_{high}, maps to the expression for surprise. The low arousal prototype, P_{low}, corresponds to the expression for fatigue (note that sleep is a behavioral response, so it is covered in a different motor subsystem). The positive valence prototype, $P_{positive}$, maps to a content expression. The negative valence prototype, $P_{negative}$, resembles an unhappy expression. The closed stance prototype, P_{closed}, resembles a stern expression, and the open stance prototype, P_{open}, resembles an accepting expression.

The three affect dimensions also map to affective postures. There are six prototype postures defined which span the space. High arousal corresponds to an erect posture with a slight upward chin. Low arousal corresponds to a slouching posture where the neck lean and head tilt are lowered. The posture remains neutral over the valence dimension. An open stance corresponds to a forward lean movement, which suggests strong interest toward the stimuli the robot is leaning toward. A closed stance corresponds to withdraw, reminiscent of shrinking away from whatever the robot is looking at.

The remaining three facial prototypes are used to strongly distinguish the expressions for disgust, anger, and fear. Recall that four of the six primary emotions are characterized by negative valence. Whereas the primary six basis postures (presented above) can generate a range of negative expressions from distress to sorrow, the expressions for intense anger (rage), intense fear (terror), and intense disgust have some uniquely distinguishing features. For instance, the prototype for disgust, $P_{disgust}$, is unique in its asymmetry (typical of this expression such as the curling of one side of the lip). The prototypes for anger, P_{anger}, and fear, P_{fear}, each have a distinct configuration for the lips (furious lips form a snarl, terrified lips form a grimace).

7.3 Interpolation Approach

Each dimension of the affect space is bounded by the minimum and maximum allowable values of $(min, max) = (-1250, 1250)$. The current net affective assessment from the emotion system defines the $[A, V, S] = (a, v, s)$ point in affect space. The specific (a, v, s) values are used to *weight* the relative motor contributions of the basis postures. Using a weighted interpolation scheme, the net emotive expression, P_{net}, is computed. The contributions are computed as follows:

$$P_{net} = C_{arousal} + C_{valence} + C_{stance} \tag{1}$$

where
P_{net} is the emotive expression computed by weighted interpolation
$C_{arousal}$ is the weighted motor contribution due to the arousal state
$C_{valence}$ is the weighted motor contribution due to the valence state
C_{stance} is the weighted motor contribution due to stance state

These contributions are specified by the equations:
$C_{arousal} = \alpha P_{high} + (1 - \alpha) P_{low}$
$C_{valence} = \beta P_{positive} + (1 - \beta) P_{negative}$
$C_{stance} = F(a, v, s, n) + (1 - \delta)(\gamma P_{open} + (1 - \gamma) P_{closed})$

where the fractional interpolation coefficients are:
$\alpha, 0 \leq \alpha \leq 1$ for arousal
$\beta, 0 \leq \beta \leq 1$ for valence
$\gamma, 0 \leq \gamma \leq 1$ for stance
$\delta, 0 \leq \delta \leq 1$ for the specialized prototype postures

such that δ and $F(A, V, S, N)$ are defined as follows:
$\delta = f_{anger}(A, V, S, N) + f_{fear}(A, V, S, N) + f_{disgust}(A, V, S, N)$

$$\begin{aligned} F(A, V, S, N) = {} & f_{anger}(A, V, S, N) \cdot P_{anger} + \\ & f_{fear}(A, V, S, N) \cdot P_{fear} + \\ & f_{disgust}(A, V, S, N) \cdot P_{disgust} \end{aligned}$$

The weighting function $f_i(A, V, S, N)$ limits the influence of each specialized prototype posture to remain local to their region of affect space. Recall, there are three specialized postures, P_i, for $i = anger$, $fear$, or $disgust$. Each is located at $(A_{P_i}, V_{P_i}, S_{P_i})$ where A_{P_i} corresponds to the arousal coordinate for posture P_i, V_{P_i} corresponds to the valence coordinate, and S_{P_i} corresponds to the stance coordinate. Given the current net affective state (a, v, s) as computed by the emotion system, one can compute the displacement from (a, v, s) to each $(A_{P_i}, V_{P_i}, S_{P_i})$. For each P_i, the weighting function $f_i(A, V, S, N)$ decays linearly with distance from $(A_{P_i}, V_{P_i}, S_{P_i})$. The weight is bounded between $0 \leq f_i(A, V, S, N) \leq 1$, where the maximum value occurs at $(A_{P_i}, V_{P_i}, S_{P_i})$. The argument N defines the radius of influence which is kept fairly small so that the contribution for each specialized prototype posture does not overlap with the others.

A range of sample expressions generated with this technique is shown in Figure 8, although the system can generate a much broader range. The procedure runs in real-time, which is critical for social interaction.

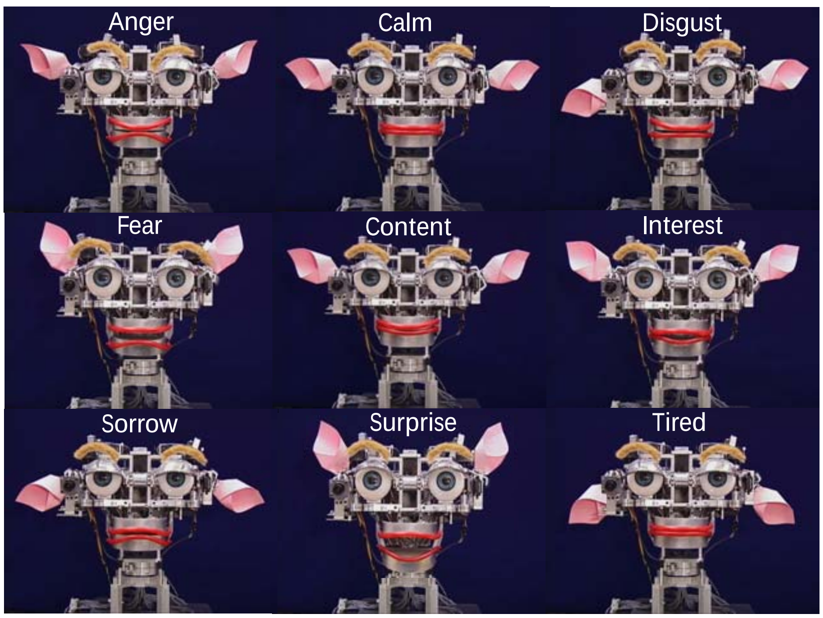

Fig. 8. Kismet is capable of generating a continuous range of expressions of various intensities by blending the basis facial postures. Facial movements correspond to affect dimensions in a principled way. A sample is shown here.

Given this three dimensional affect space, this approach resonates well with the work of Smith & Scott (1997). They posit a three dimensional space of *pleasure-displeasure* (maps to valence here), *attentional activity* (maps to arousal here), and *personal agency, control* (roughly maps to stance here). Table 1 summarizes their proposed mapping of facial actions to these dimensions. They posit a fourth dimension that relates to the intensity of the expression. For Kismet, the expressions become more intense as the affect state moves to more extreme values in the affect space. As positive valence increases, Kismet's lips turn upward, the mouth opens, and the eyebrows relax. However, as valence decreases, the brows furrow, the jaw closes, and the lips turn downward. Along the arousal dimension, the ears perk, the eyes widen, and the mouth opens as arousal increases. Along the stance dimension, increasing positive values cause the eyebrows to arc outwards, the mouth to open, the ears to open, and the eyes to widen. These face actions roughly correspond to a decrease in personal agency/control in Smith and Scott's framework. For Kismet, it engenders an expression that looks more eager and accepting (or more uncertain for negative emotions). Although Kismet's dimensions do not map exactly to those hypothesized by Smith & Scott (1997), the idea of combining meaningful face action units in a principled manner to span the space of facial expressions, and to also relate them in a consistent way to emotion categories, holds strong.

8 Static Evaluation of Expressive Behavior

To explore how recognizable Kismet's facial expressions are to people, a questionnaire was devised. Given the wide variation in language that people use to describe expressions and the small number of subjects, a forced choice paradigm was adopted.

Seventeen subjects filled out the questionnaire. Most of the subjects were children 12 years of age (note that the ability to recognize expressions continues to develop, reaching adult level competence at approximately 14 years of age (Kolb et al., 1992)). There were six girls, six boys, three adult men, and two adult women. None of the adults had seen the robot before. Some of the children reported minimal familiarity through reading a children's magazine article. There were seven pages in the questionnaire. Each page had a large color image of Kismet displaying one of seven expressions (i.e., for anger, disgust, fear, happiness, sorrow, surprise, and a stern expression). The subjects could choose the best match from ten possible labels (i.e., *accepting, anger, bored, disgust, fear, joy, interest, sorrow, stern, surprise*). In a follow-up question, they could circle any other labels that they thought could also apply. With respect to their best-choice answer, they were asked to specify on a ten-point scale how confident they were of their answer, and how intense they found the expression. The compiled results are shown in Table 3. The subjects' responses were significantly above random choice (10 percent), ranging from 47 percent to 83 percent.

Kismet's surprise expression seems to convey positive valence, as some subjects matched it to "joy." The knitting of the brow in Kismet's stern expression is most likely responsible for the associations with negative emotions such as anger and sorrow. Often, negatively valenced expressions were misclassified with negatively valenced labels. For instance, labeling the sad expression with "fear," or the disgust expression with "anger" or "fear." Kismet's expression for fear seems to give people the most difficulty. The lip mechanics probably

	accepting	anger	bored	disgust	fear	joy	interest	sorrow	stern	surprise	% correct
anger	5.9	76.5	0	0	5.9	11.7	0	0	0	0	76.5
disgust	0	17.6	0	70.6	5.9	0	0	0	5.9	0	70.6
fear	5.9	5.9	0	0	47.1	17.6	5.9	0	0	17.6	47.1
joy	11.7	0	5.9	0	0	82.4	0	0	0	0	82.4
sorrow	0	5.9	0	0	11.7	0	0	83.4	0	0	83.4
stern	7.7	15.4	0	7.7	0	0	0	15.4	53.8	0	53.8
surprise	0	0	0	0	0	17.6	0	0	0	82.4	82.4

Forced-Choice Percentage (random=10%)

Table 3
This table summarizes the results of the color-image-based evaluation. The questionnaire was forced choice where the subject chose the emotive word that best matched the picture. Each row label corresponds to an image of Kismet showing one of seven possible expressions. Each of the column labels corresponds to one of ten emotion words the subject could match to the Kismet image.

	anger	disgust	fear	joy	interest	sorrow	urprise	% correct
anger	86	0	0	14	0	0	0	86
disgust	0	86	0	0	0	14	0	86
fear	0	0	86	0	0	0	14	86
joy	0	0	0	57	28	0	15	57
interest	0	0	0	0	71	0	29	71
sorrow	14	0	0	0	0	86	0	86
surprise	0	0	29	0	0	0	71	71

Forced-Choice Percentage (random=14%)

Table 4
This table summarizes the results of the video evaluation.

account for the association with "joy," where the ends of the lips curve up a bit at interface junction with the motors. The wide eyes, elevated brows, and elevated ears suggest high arousal. This may account for the confusion with "surprise."

The still image studies were useful in understanding how people read Kismet's facial expressions, but it says very little about expressive posturing. Humans and animals not only express with their face, but with their entire body. To explore this issue for Kismet, we showed a small group of subjects a set of video clips.

There were seven people who filled out a second questionnaire. Six were children of age 12, four boys and two girls. One was an adult female. In each clip Kismet performs a coordinated expression using face and body posture. There were seven videos for the expressions of anger, disgust, fear, joy, interest, sorrow, and surprise. Using a forced-choice paradigm, for each video the subject was asked to select a word that best described the robot's expression (*anger, disgust, fear, joy, interest, sorrow*, or *surprise*). On a ten-point scale, the subjects were also asked to rate the intensity of the robot's expression and the certainty of their answer. They were also asked to write down any comments they had. The results are compiled in Table 4. Random chance is 14 percent.

The subjects performed significantly above chance, with overall stronger recognition performance than on the still images alone. The video segments of "anger," "disgust," "fear," and "sorrow" were correctly classified with a higher percentage than the still images. However, there were substantially fewer subjects who participated in the video evaluation than the still image evaluation. The recognition of "joy" most likely dipped from the still-image counterpart because it was sometimes confused with the expression of interest in the video study. The perked ears, attentive eyes, and smile give the robot a sense of expectation that could be interpreted as interest.

Misclassifications are strongly correlated with expressions having similar facial or postural components. "Surprise" was sometimes confused for "fear;" both have a quick withdraw postural shift (the fearful withdraw is more of a cowering movement whereas the surprise posture has more of an erect quality) with wide eyes and elevated ears. "Surprise" was sometimes confused with "interest" as well. Both have an alert and attentive quality, but interest is an approaching movement whereas surprise is more of a startled movement. "Sorrow" was sometimes confused with "disgust;" both are negative expressions with a downward component to the posture. The sorrow posture shift is more down and "sagging," whereas the disgust posture is a slow "shrinking" retreat.

Overall, the data gathered from these small evaluations suggest that people with little to no familiarity with the robot are able to interpret the robot's facial expressions and affective posturing. For this data set, there was no clear distinction in recognition performance between adults versus children, or males versus females. They map the expressions to corresponding emotion labels with reasonable consistency, and many of the errors can be explained through similarity in facial features or similarity in affective assessment (e.g., shared aspects of arousal or valence).

The data from the video studies suggest that witnessing the movement of the robot's face and body strengthens the recognition of the expression. The average recognition of emotional expressions for the static images is 70.9% verses an average of 77.6% for the video case. This compares favorably to the results reported in Canamero & Fredslund (2001) with an average recognition rate of 55% for adults and 48% for children given a similar multiple choice test.

9 Evaluation of Real-Time Affective Interactions with People

The studies presented in the previous section are static; they do not involve a human interacting with Kismet in real-time. This section summarizes the dynamic quality of interaction that transpires between people and Kismet.

The design of Kismet's emotion system enables the robot to use expressive social cues to tune the human's behavior so that both perform well during the interaction. Kismet's motivation system is explicitly designed so that a state of "well-being" for the robot corresponds to an environment that affords a high learning potential. As reported in Breazeal (1998) this often maps to having a person actively engaging the robot in a manner that is neither under-stimulating nor overwhelming. Furthermore, the robot actively regulates the relationship between itself and its environment, to bring itself into contact with desired stimuli and to avoid undesired stimuli (Breazeal & Scassellati, 2000).

All the while, the cognitive appraisals leading to these actions are displayed on the robot's face, to which people read and respond to accordingly. For instance, if the robot appears bored (resulting from low arousal), a person might respond by trying to engage Kismet with a toy. If the person waves the toy too fast and too close to its face, the affective state swiftly but smoothly transitions to become more aroused, more negative, and more withdrawn. As a result, Kismet looks increasingly distressed. People typically realize that the robot is reacting adversely to the situation and back off. If they persist, however, the expression turns fearful and an escape response (to protect the robot) ensues. Interactions such as these are reported in Breazeal & Scassellati (2000).

In Breazeal & Aryananda (2002) and Breazeal (2002*b*) we report findings from another set of interaction studies that explored the communication of affective intent to Kismet through tone of voice. In this set of studies, Kismet recognized four affective intents (i.e., praise, prohibition, attentional bids, and soothing) from a person's vocal prosody. A recognizer was designed to categorize these four manners of speech, each of them having an associated releaser mechanism. When interfaced with Kismet's emotion system, the person could manipulate the robot's affective state through tone of voice, causing the robot to become more positive through praising tones, more aroused through alerting tones, more "sad" through scolding tones, and moderately aroused through soothing tones.

In this study, five female subjects (ranging from 23 to 54 years old) were asked to interact with Kismet in different languages (English, Russian, French, German, and Indonesian). One of the subjects had interacted with Kismet frequently in the past, and spoke to the robot in either English or Indonesian for this experiment. Subjects were instructed to express each affective intent (approval, attention, prohibition, and soothing) and signal when they felt that they had communicated it to the robot.

Recorded events show that subjects in the study made ready use of Kismet's expressive feedback to assess when the robot "understood" them. The robot's expressive repertoire included both facial expressions and shifts in body posture. The subjects varied in their sensitivity to the robot's expressive feedback, but all used the robot's expressive cues 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 expression. If the wrong expression appeared, they often used strongly exaggerated their prosody to correct the "misunderstanding."

Kismet's expression through face and body posture becomes more intense as the activation level of the corresponding emotion process increases. 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." The subjects could discern these intensity differences, and several modulated their speech to influence them. For example, in one trial a subject scolded Kismet, to which it dipped its head. However, the subject continued to prohibit Kismet with a lower and lower voice until Kismet eventually frowned. Only then did the subject stop her prohibitions.

During course of the interaction, several interesting dynamic social phenomena arose. Often these occurred in the context of prohibiting the robot. For instance, several of the subjects reported experiencing a very strong emotional response immediately after "successfully" prohibiting the robot. 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." Subjects were often very apologetic throughout their prohibition session. 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 we 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.

These interaction studies showcase one of the significant contributions of the design of Kismet's emotion and expression system. Namely, the ability to engage people in face-to-face, rich, dynamic, mutually regulated, and closely coupled affective interactions. The resulting interactions are quite engaging because the robot's expressive behavior is timely and appropriately synchronized with the human's behavior at fine-grained time scales (i.e., less than a second). This attention to temporal detail and its synchrony with real-time human behavior is critical in order to establish a natural flow and rhythm to the human-robot interaction that is characteristic of human-human interaction. As a result, the interaction is not only stimulating for the robot, but it is compelling for the person who interacts with it. Kismet's architecture affords these temporal characteristics because of its mechanistic, circuit-like approach (in contrast to a symbolic rule-based approach) where time is explicitly represented in the processes themselves and in how they dynamically interact with each other (e.g., decay rates, minimum activation time, habituation time, etc.). In short, to offer a high quality (i.e., compelling and engaging) interaction with humans, it is important that the robot not only do the right thing, but also at the right time and in the right manner.

10 Discussion and Summary

Our implementation of basic emotive responses takes its inspiration from ethological models for the organization of behavior in animals, and how this relates to the concept of drives and triggering events (i.e., internal releasing mechanisms) (Tinbergen, 1951; Lorenz, 1973). As such, it is similar in spirit to the *Cathexis* system of Velasquez (1996). However, there are significant differences in their respective designs given the emphasis our system places on social interaction with humans, whereas Velasquez's robotic instantiations are more concerned with "survival" related issues.

First, our system explicitly represents the robot's affective assessment in terms of a lower dimensional space (i.e., *arousal*, *valence*, and *stance*) that spans the relationship between different "emotional" states and their expression (as denoted in the affect space). The $[A, V, S]$ trio serve as a common currency that allows for a unified way to integrate information from disparate souces including behavior, drives, perception, and emotion. In addition, this representation allows us to take a unified approach with respect to the assessment of eliciting conditions, the representation of "emotional" state, and the generation of facial expression. The Cathexis system does not offer a representation that allows all these different aspects to be handled in a unified way.

Second, whereas Cathexis models emotion, it does not present a methodology for generating expression. Expression, however, plays a critical role in social interaction with people. Kismet's face serves as a window into its underlying affective assessment. This is the reason why $[A, V, S]$ is explicitly represented in our system (by adapting Damasio's Somatic Marker Hypothesis to tag the releasers). There is information contained not only in the expression itself, but also in how this expression changes over time (i.e., the trajectory through affect space). As discussed in Section 9, people read both kinds of information in Kismet's expressive feedback, and actively use them to influence its "emotional" state. Furthermore, because the arousal, valance, and stance values are never completely static, the robot's face has a compelling animated quality—undergoing subtle shifts while in a given "emotional" state, and undergoing more dramatic shifts when changing to a different "emotional" state. This significantly contributes to the life-like quality of the robot's expression and makes the robot quite engaging for the human.

Third, the Cathexis system emphasizes *emotion based control* where the emotion system plays the dominant role in behavior arbitration (Velasquez, 1998). Hence, the vast majority of behaviors are evoked as part of an emotional response. This benefits the robot's "survival" where it must avoid potentially dangerous situations and seek out "food" sources. In contrast, Kismet operates in a benevolent, social world. Most of its time is spent in a few "emotional" states (e.g., "content" or "interested") when interacting well with an engaged human. Hence, the behavior system (rather than the emotion system) plays the dominant role in behavior arbitration— considering both external perceptual events and a variety of internal events (of which `emotion` is one of several). This difference in control strategy allows the robot to engage in a wide assortment of play-related behaviors with a person when in a "contented" or "interested" state. The emotion system exerts its influence on control in order to bring the robot back into a "contented" state through behavioral homeostasis when conditions drift away from what is desirable (such as either a lack of engagement, or too much engagement). A full description of the behavior system is beyond the scope of this paper, but can be found in Breazeal (2002*a*).

Kismet's expressive interactions with people are dynamic and closely coupled to how people are engaging the robot, and thereby influencing its internal affective state as governed by its emotional model. There are several advantages to generating the robot's facial expression from its affective state.

First, this technique allows the robot's facial expression to reflect the nuance of the underlying assessment (i.e., arousal, valence, and stance). Even though there is a discrete number of emotion processes, the expressive behavior spans a continuous space allowing the robot to blend its expressions to create nuanced versions such as a sly grin (i.e., a smile with knitted brows, squinted eyelids, and ears folded back), or a stern look (e.g., knitted brows and folded ears with an otherwise neutral face), and more. Recall from section 4.2 a componential approach was adopted because it allows for the mixing of facial components to convey a wide range of affective messages beyond those canonical configurations of the basic categories (Smith & Scott, 1997). It is difficult to see how a category-based approach to expression would accommodate this degree of nuance and flexibility.

Second, a componential approach lends clarity to the facial expression since the robot can only be in a single affective state at a time (by 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 and informative. By having the face mirror this trajectory, the observer has immediate feedback as to how their behavior is influencing the robot's internal state (i.e., how the robot is affectively assessing the current situation). 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 (as described in Breazeal & Aryananda (2002)) 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. However, if the face appeared to grow more aroused, then the person may stop trying to comfort the robot verbally and perhaps try to please the robot by showing it a colorful toy.

Although the challenge of building robots that interact with people may share some issues with the design of computer interfaces, robots and computers are profoundly different technologies in important ways. Specifically, robots not only have to carry out their tasks, they also have to survive in the human environment. From the robot's perspective, the real world is complex, unpredictable, partially knowable, and continually changing. The ability of robots to adapt and learn in such an environment is fundamental. For robots, social and emotive qualities serve not only to "lubricate" the interface between humans and robots, but also play a pragmatic role in promoting survival, self maintenance, learning, decision making, attention, and more (Velasquez, 1997; Canamero, 1997; Yoon et al., 2000). Hence, when designing robots that interact with humans in the real world, the issue is not so much whether robots should have social and emotive characteristics, but of what kind (Sloman, 1981).

The purpose of this work is to investigate the role that emotion and and expression play in social, face-to-face interaction between robots and people. Kismet engages people not only to communicate and interact with them, but to also promote its "well being". The interactions that transpire between robot and human are in the spirit of those shared between adults and pre-verbal infants. Hence, the interactions are fundamentally physical, affective, and follow the temporal dynamics of an expressive proto-dialog (Breazeal & Scassellati, 1999). This type of interaction places special demands on the robot's architecture that we have identified and addressed in this work, i.e., the use of basic emotion models and behavioral homeostasis to implement proto-social responses, the readability of the robot's facial expressions, and the temporal characteristics of affective exchanges between robot and human.

Whereas past work into computational models of emotion often views their expressive component as being of secondary importance, it is of critical importance when interacting socially with humans. The robot's observable behavior and the manner in which it responds and reacts to people profoundly shapes the interaction and the mental model people have for the robot (please refer back to section 4). In numerous examples during our interaction experiments (Breazeal & Aryananda, 2002; Breazeal, 2002*b*), people treat Kismet as a social creature. They interpret Kismet's behavior as the product of intents, beliefs, desires, and feelings, and respond to Kismet in these terms. Their behavior is closely and contingently synchronized with that of the robot, and vice versa.

As we have stressed earlier, it is important that the robot not only do the right thing, but also at the right time and in the right manner. To achieve this, the emotion system and the expressive motor system must work in concert. To facilitate this, we adopted a common representation of affective state (i.e., the three dimensions, $[A, V, S]$) so that we could take a unified approach with respect to the assessment of eliciting conditions, the representation of "emotional" state, and the generation of facial expression. We used somatic markers to convert the myriad of factors that influence the robot's affective state (e.g., behaviors, drives, perceptions, etc.) into this common currency, $[A, V, S]$.

11 Future Work

The ultimate scientific question we want to address with this kind of work is to understand the role emotion-like processes might play in socially situated learning and social development of robots that co-exist with people in the human environment. We adopt an approach of synthesis and iteration where the goal is to build a robot that is capable of learning from people from social interactions, and learns to be more socially sophisticated in the process. This goal has guided Kismet's design, much of it was inspired by insights from the field of Developmental Psychology.

As with human infants, we believe that treating Kismet as a fully socially responsive creature (before it is genuinely so) will be critical for its own social development (Bullowa, 1979). The emotive and other proto-social responses of human infants play an important role in this process, as we believe they do for Kismet. They launch the robot into valuable social interactions that afford great learning potential (it is very common for a person to try to teach Kismet something while playing with it). Furthermore, a person's behavior is made more consistent and predictable for the robot if he/she treats Kismet as a socially responsive creature, believing that it shares the same meanings that he/she applies to the interaction. This allows routine and predictable sequences to be established, that the robot could eventually exploit to learn the significance its actions and expressions have for others (i.e., learning shared meanings). Furthermore, through such interactions, the robot could discover what sorts of activity on its part will get particular responses from those who interact with it (i.e., learning the pragmatics of use).

This is the long term vision, but there are a few logical next steps. As presented in this paper, Kismet is able to engage in social and affective interactions that afford rich learning potential. One important next step is for Kismet to learn from these exchanges, while exploring the role of "emotion" in this process. Social referencing is a familiar example—human infants often read the emotive expression of the caregiver when confronted by a novel situation. The infant, not knowing how to assess the situation, applies the adult's affective assessment to the situation and behaves accordingly (often learning to avoid a potentially dangerous situation if the caregiver exhibits anxiety or fear).

Another important next step is for Kismet to acquire its own mental models of people. Currently, Kismet does not reason about the emotional state of others. The ability to recognize, understand, and reason about another's emotional state is an important ability for having a theory of mind about other people, which is considered by many to be a requisite of adult-level social intelligence (Dennett, 1987). There are a few systems that have been designed to reason about human emotions, typically based on symbolic models that analyze text input from a human user (Ortony et al., 1988; Elliot, 1992; Reilly, 1996). However, this has yet to be demonstrated by a robot that engages people in face-to-face interaction. Scassellati (2000) presents an early exploration into endowing humanoid robots with a theory of mind, but focuses on the ability to establish shared attention rather than exploring emotive aspects.

The work described in this paper is an early step towards these questions. It is our hope that the insights we glean from this exploration might also contribute to a deeper understanding of how human infants learn from people, how they develop socially, and the role of emotion in these processes.

Acknowledgments

The author gratefully acknowledges the creativity and ingenuity of the members of the Humanoid Robotics Group at the MIT Artificial Intelligence Lab. In particular, Lijin Aryananda played a critical role in the recognition of affective intent in robot directed speech. This work was funded by NTT and DARPA contract DABT 63-99-1-0012. The author is funded by the Things that Think and Digital Life consortia sponsors of the MIT Media Lab.

References

Atkeson, C. & Schaal, S. (1997*a*), Learning tasks from single demonstration, *in* 'Proceedings of the International Conference on Robotics and Automation (ICRA97)', IEEE, pp. 1706–1712.

Atkeson, C. & Schaal, S. (1997*b*), Robot learning from demonstration, *in* 'Proceedings of the International Conference on Machine Learning (ICML97)', Morgan Kaufman, San Francisco, CA, pp. 12–20.

Billard, A. (2002), Imitation: a means to enhance learning of a synthetic protolanguage in an autonomous robot, *in* K. Dautenhahn & C. Nehaniv, eds, 'Imitation in Animals and Artifacts', MIT Press, pp. 281–310.

Breazeal, C. (1998), A motivational system for regulating human-robot interaction, *in* 'Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI98)', Madison, WI, pp. 54–61.

Breazeal, C. (2002*a*), *Designing Sociable Robots*, MIT Press, Cambridge, MA.

Breazeal, C. (2002*b*), 'Regulation and Entrainment in Human-Robot Interaction', *International Journal of Robotics Research*. forthcoming.

Breazeal, C. & Aryananda, L. (2002), 'Recognition of affective communicative intent in robot-directed speech', *Autonomous Robots* **12**(1), 83–104.

Breazeal, C. & Scassellati, B. (1999), How to build robots that make friends and influence people, *in* 'Proceedings of the 1999 IEEE/RSJ International

Conference on Intelligent Robots and Systems (IROS99)', Kyonju, Korea, pp. 858–863.

Breazeal, C. & Scassellati, B. (2000), 'Infant-like social interactions between a robot and a human caretaker', *Adaptive Behavior* **8**(1), 47–72.

Brooks, R., Breazeal, C., Marjanovic, M., Scassellati, B. & Williamson, M. (1999), The Cog Project: Building a humanoid robot, *in* C. L. Nehaniv, ed., 'Computation for Metaphors, Analogy and Agents', Vol. 1562 of *Springer Lecture Notes in Artificial Intelligence*, Springer-Verlag, New York, NY.

Bruce, A., Nourbakshs, I. & Simmons, R. (2001), The role of expressivness and attention in human-robot interaction, *in* 'Proceedings of the 2001 AAAI Fall Symposium'.

Bullowa, M. (1979), *Before Speech: The Beginning of Interpersonal Communication*, Cambridge University Press, Cambridge, UK.

Burgard, W., Cremers, A., Fox, D., Haehnel, D., Lakemeyer, G., Schulz, D., Steiner, W. & Thrun, S. (1998), The interactive museum tour-guide robot, *in* 'Proceedings of the Fifthteenth National Conference on Artificial Intelligence (AAAI98)', Madison, WI, pp. 11–18.

Canamero, D. (1997), Modeling motivations and emotions as a basis for intelligent behavior, *in* L. Johnson, ed., 'Proceedings of the First International Conference on Autonomous Agents (Agents97)', ACM Press, pp. 148–155.

Canamero, L. & Fredslund, J. (2001), 'I show how I like you: Can you read my face?', *IEEE Systems, Man, and Cybernetics – Part A* **31**(5), 454–459.

Cassell, J. (1999), Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents, *in* J. Cassell, J. Sullivan, S. Prevost & E. Churchill, eds, 'Embodied Conversational Agents', MIT Press, Cambridge, MA, pp. 1–27.

Cassell, J., Bickmore, T., Campbell, L., Vilhjalmsson, H. & Yan, H. (2000), Human conversation as a system framework: Designing embodied conversation agents, *in* J. Cassell, J. Sullivan, S. Prevost & E. Churchill, eds, 'Embodied Conversational Agents', MIT Press, Cambridge, MA, pp. 29–63.

Damasio, A. (1994), *Descartes Error: Emotion, Reason, and the Human Brain*, G.P. Putnam's Sons, New York, NY.

Dario, P. & Susani, G. (1996), Physical and psychological interactions between humans and robots in the home environment, *in* 'Proceedings of the First International Symposium on Humanoid Robots (HURO96)', Tokyo, Japan, pp. 5–16.

Darwin, C. (1872), *The Expression of the Emotions in Man and Animals*, John Murray, London, UK.

Demiris, J. & Hayes, G. (2002), Imitation as a dual-route process featuring predictive and learning components: A biologically plausible computational model, *in* K. Dautenhahn & C. Nehaniv, eds, 'Imitation in Animals and Artifacts', MIT Press, pp. 321–361.

Dennett, D. (1987), *The Intentional Stance*, MIT Press, Cambridge, MA.

Ekman, P. (1992), 'Are there basic emotions?', *Psychological Review* **99**(3), 550–553.

Ekman, P. & Friesen, W. (1982), Measuring facial movement with the Facial Action Coding System, *in* 'Emotion in the Human Face', Cambridge University Press, Cambridge, UK, pp. 178–211.

Ekman, P. & Oster, H. (1982), Review of research, 1970 to 1980, *in* P. Ekman, ed., 'Emotion in the Human Face', Cambridge University Press, Cambridge, UK, pp. 147–174.

Elliot, C. D. (1992), The Affective Reasoner: A Process Model of Emotions in a Multi-Agent System, PhD thesis, Northwestern University, Institute for the Learning Sciences, Chicago, IL.

Frijda, N. (1994*a*), Emotions are functional, most of the time, *in* P. Ekman & R. Davidson, eds, 'The Nature of Emotion', Oxford University Press, New York, NY, pp. 112–122.

Frijda, N. (1994*b*), Emotions require cognitions, even if simple ones, *in* P. Ekman & R. Davidson, eds, 'The Nature of Emotion', Oxford University Press, New York, NY, pp. 197–202.

Frijda, N. (1994*c*), Universal antecedents exist, and are interesting, *in* P. Ekman & R. Davidson, eds, 'The Nature of Emotion', Oxford University Press, New York, NY, pp. 146–149.

Hara, F. (1998), Personality characterization of animate face robot through interactive communication with human, *in* 'Proceedings of the 1998 International Advanced Robotics Program (IARP98)', Tsukuba, Japan, pp. IV–1.

Hara, F. & Kobayashi, H. (1996), A face robot able to recognize and produce facial expression, *in* 'Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS96)', pp. 1600–1607.

Hirai, K. (1998), Humanoid robot and its applications, *in* 'Proceedings of the 1998 International Advanced Robot Program (IARP98)', Tsukuba, Japan, pp. V–1.

Izard, C. (1977), *Human Emotions*, Plenum Press, New York, NY.

Izard, C. (1993), 'Four systems for emotion activation: Cognitive and noncognitive processes', *Psychological Review* **100**, 68–90.

Izard, C. (1994), Cognition is one of four types of emotion activating systems, *in* P. Ekman & R. Davidson, eds, 'The Nature of Emotion', Oxford University Press, New York, NY, pp. 203–208.

Kawamura, K., Wilkes, D., Pack, T., Bishay, M. & Barile, J. (1996), Humanoids: future robots for home and factory, *in* 'Proceedings of the First International Symposium on Humanoid Robots (HURO96)', Tokyo, Japan, pp. 53–62.

Kiesler, S. & Goetz, J. (2002), Mental models of robotic assistants, *in* 'Proceedings of the Conference on Human Fators in Computing Systems (CHI2002)', Minneapolis, MN, pp. 576–577.

Kolb, B., Wilson, B. & Laughlin, T. (1992), 'Developmental changes in the recognition and comprehension of facial expression: Implications for frontal lobe function', *Brain and Cognition* pp. 74–84.

Lazarus, R. (1991), *Emotion and Adaptation*, Oxford University Press, New York, NY.

Lazarus, R. (1994), Universal antecedents of the emotions, *in* P. Ekman & R. Davidson, eds, 'The Nature of Emotion', Oxford University Press, New York, NY, pp. 163–171.

Levenson, R. (1994), Human emotions: A functional view, *in* P. Ekman & R. Davidson, eds, 'The Nature of Emotion', Oxford University Press, New York, NY, pp. 123–126.

Lorenz, K. (1973), *Foundations of Ethology*, Springer-Verlag, New York, NY.

Mataric, M. (2000), 'Getting humanoids to move and imitate', *IEEE Intelligent Systems* pp. 18–23.

McFarland, D. & Bosser, T. (1993), *Intelligent Behavior in Animals and Robots*, MIT Press, Cambridge, MA.

Nourbakhsh, I., Bobenage, J., Grange, S., Lutz, R., Meyer, R. & Soto, A. (1999), 'An affective mobile educator with a full-time job', *Artificial Intelligence* **114**(1–2), 95–124.

Ortony, A., Clore, G. & Collins, A. (1988), *The Cognitive Structure of Emotion*, Cambridge University Press, Cambridge, UK.

Plutchik, R. (1984), Emotions: A general psychevolutionary theory, *in* K. Scherer & P. Elkman, eds, 'Approaches to Emotion', Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 197–219.

Plutchik, R. (1991), *The Emotions*, University Press of America, Lanham, MD.

Reeves, B. & Nass, C. (1996), *The Media Equation*, CSLI Publications, Stanford, CA.

Reilly, S. (1996), Believable Social and Emotional Agents, PhD thesis, Carnegie Mellon University, School of Computer Science, Pittsburgh, PA.

Rickel, J. & Johnson, W. L. (2000), Task-oriented collaboration with embodied agents in virtual worlds, *in* J. Cassell, J. Sullivan, S. Prevost & E. Churchill, eds, 'Embodied Conversational Agents', MIT Press, Cambridge, MA, pp. 95–122.

Russell, J. (1997), Reading emotions from and into faces: Resurrecting a dimensional-contextual perspective, *in* J. Russell & J. Fernandez-Dols, eds, 'The psychology of Facial Expression', Cambridge university press, Cambridge, UK, pp. 295–320.

Scassellati, B. (2000), Foundations for a Theory of Mind for a Humanoid Robot, PhD thesis, MIT Department of Electrical Engineering and Computer Science.

Schaal, S. (1997), Learning from demonstration, *in* 'Proceedings of the 1997 Conference on Neural Information Processing Systems (NIPS97)', Denver, CO, pp. 1040–1046.

Scheef, M., Pinto, J., Rahardja, K., Snibbe, S. & Tow, R. (2000), Experiences with Sparky, a social robot, *in* 'Proceedings of the 2000 Workshop on Interactive Robot Entertainment', Pittsburgh, PA.

Scherer, K. (1984), On the nature and function of emotion: A component process approach, *in* K. Scherer & P. Ekman, eds, 'Approaches to Emotion', Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 293–317.

Scherer, K. (1994), Evidence for both universality and cultural specificity of emotion elicitation, *in* P. Ekman & R. Davidson, eds, 'The Nature of Emotion', Oxford University Press, New York, NY, pp. 172–175.

Sloman, A. (1981), Why robots will have emotions, *in* 'Proceedings of IJCAI81'.

Smith, C. (1989), 'Dimensions of appraisal and physiological response in emotion', *Journal of Personality and Social Psychology* **56**, 339–353.

Smith, C. & Scott, H. (1997), A componential approach to the meaning of facial expressions, *in* J. Russell & J. Fernandez-Dols, eds, 'The Psychology of Facial Expression', Cambridge University Press, Cambridge, UK, pp. 229–254.

Takanobu, H., Takanishi, A., Hirano, S., Kato, I., Sato, K. & Umetsu, T. (1999), Development of humanoid robot heads for natural human-robot communication, *in* 'Proceedings of Second International Symposium on Humanoid Robots (HURO99)', Tokyo, Japan, pp. 21–28.

Takeuchi, A. & Nagao, K. (1993), Communicative facial displays as a new conversational modality, *in* 'Proceedings of the 1993 ACM Conference on Human Factors in Computing Systems (ACM SIGCHI93)', Amsterdam, The Netherlands, pp. 187–193.

Tinbergen, N. (1951), *The Study of Instinct*, Oxford University Press, New York, NY.

Velasquez, J. (1996), Cathexis, A Computational Model for the Generation of Emotions and their Influence in the Behavior of Autonomous Agents, Master's thesis, MIT.

Velasquez, J. (1997), Modeling emotions and other motivations in synthetic agents, *in* 'Proceedings of the 1997 National Conference on Artificial Intelligence (AAAI97)', Rrovidence, RI, pp. 10–15.

Velasquez, J. (1998), A computational framework for emotion-based control, *in* 'Proceedings of the SAB98 Workshop on Grounding Emotions in Adaptive Systems'.

Vilhjalmsson, H. & Cassell, J. (1998), BodyChat: Autonomous communicative behaviors in avatars, *in* 'Proceedings of the Second Annual Conference on Automous Agents (Agents98)', Minneapolis, MN, pp. 269–276.

Woodworth, R. (1938), *Experimental Psychology*, Holt, New York.

Yoon, S., Blumberg, B. & Schneider, G. (2000), Motivation driven learning for interactive synthetic characters, *in* 'Proceedings of the Fourth International Conference on Autonomous Agents (Agents00)', Barcelona, Spain.