Time-Varying Affective Response for Humanoid Robots[*]

Lilia Moshkina[1], Ronald C. Arkin[1], Jamee K. Lee[2], and HyunRyong Jung[2]

[1] Georgia Tech Mobile Robot Laboratory, Atlanta, GA, USA 30332
{lilia,arkin}@cc.gatech.edu
[2] Samsung Advanced Institute of Technology, Kiheung, South Korea
{jamee.lee,hyunryong.jung}@samsung.com

Abstract. This paper describes the design of a complex time-varying affective architecture. It is an expansion of the TAME architecture (traits, attitudes, moods, and emotions) as applied to humanoid robotics. It particular it is intended to promote effective human-robot interaction by conveying the robot's affective state to the user in an easy-to-interpret manner.

Keywords: Humanoids, emotions, affective phenomena, robot architectures.

1 Introduction

Based on our considerable experience implementing affective phenomena in robotic systems (see [1] for a summary), we are now considering the application of sophisticated cognitive models of human Traits, Attitudes, Moods, and Emotions (TAME). These affective states are embedded into a novel architecture and designed to influence the perception of a user regarding the robot's internal state and the human-robot relationship itself. Recent work by Arkin et al in non-verbal communication [2] and emotional state for the AIBO [3] addressed powerful yet less complex means for accomplishing these tasks. Introducing time-varying affective states that range over multiple time scales spanning from an agent's lifetime to mere seconds with direction towards specific objects or the world in general provides the power to generate heretofore unobtained richness of affective expression. This paper describes the cognitive underpinnings of this work in the context of humanoid robots and presents the directions being taken in this recently initiated project to implement it upon a small humanoid robot.

2 Related Work

Although most work on humanoids focuses on the physical aspects (e.g., perfecting walking gaits, sensors or appearance), there are some who add affect into the mix. For example, humanoid Waseda Eye No. 4 Refined [4] combines emotions, moods, and personality. The overall goal of the system is to achieve smooth and effective communication for a humanoid robot. Although many elements of this system are not

psychologically or biologically founded, it provides a few interesting mechanisms, such as modeling personality's influence on emotion via a variety of coefficient matrices and using internal-clock activation component in moods.

Fukuda et al. [5] also include the notions of emotions and moods in their Character Robot Face; emotions are represented as semantic networks, and the combination of currently active emotions is deemed as mood. Two other humanoid robotic head robots, Kismet [6] and MEXI [7] have emotion and drive systems. Kismet is modeled after an infant, and is capable of proto-social responses, including emotional expressions, which are based on its affective state. In MEXI, the Emotion Engine is composed of a set of basic emotions (positive that it strives to achieve and negative it tries to avoid) and homeostatic drives. Space prevents a more complete description of other related projects.

3   Cognitive Basis of TAME

The TAME affective architecture has been initially tested on the entertainment robot dog Aibo [8], but its application to a humanoid robot is fairly straightforward in principle. In fact, using the framework for a humanoid will provide a number of advantages. The synergistic combination of affective phenomena focuses on long-term, and sometimes subtle, effect on robotic behavior, which fits well with one of the main goals for creating human-like robots - making the communication between them more natural, commonplace and prolonged, where machines act as partners rather than bystanders. The second advantage of applying TAME to humanoids is their expressive potential, exhibited not only in facial and bodily expressions (e.g., a smile, a shoulder shrug, a handshake), but also in a variety of tasks they could perform for which human-like personalities are readily applicable. The framework itself takes inspiration from a large number of theories and findings from Personality, Emotion, Mood and Attitude areas of Psychology, which are adapted to enhance robotic behavior.

3.1   Overview

The Affective Module, the core of TAME, is subdivided into Dispositions and Affective State. Dispositions include personality Traits and affective Attitudes, and represent a propensity to behave in a certain way; they are more or less persistent, long-lasting, and either slowly changing (attitudes) or permanent (traits) throughout robot's "life". Affective State consists of Emotions and Moods, more fleeting and transient affects, manifesting as either high-intensity, short duration peaks (emotions) or slow smooth undulations (moods). Table 1 summarizes differences in duration and temporal changes of these four components.

Another direction along which these components differ is object specificity: emotions and attitudes appear and change in response to particular stimuli (such as fear in the presence of an attacker or dislike towards an unfriendly person), whereas traits and moods are diffuse and not object-specific – they manifest regardless of presence or absence of objects. Although they all can be categorized differently and each have a distinct function and purpose, we cannot regard these phenomena as independent, as they strongly influence each other and interweave to create a greater illusion of life.

The Affective Module fits within behavior-based robotic control [9] by first processing relevant perceptual input (be it color and distance to certain emotion-eliciting objects or level of light affecting moods) and then influencing behavioral parameters of affected low-level behaviors and/or the behavior coordination gains as they are comprised into behavioral assemblages (see Figure 1).

Table 1. Summary of Time-varying Aspects of TAME Components

|	Traits	Attitudes	Moods	Emotions
Duration	Life-long	A few days to a few years	A few hours to a few weeks	A few seconds to a few minutes
Change in Time	Time-invariant	Persistent across time; change slowly with the number of times an object of attitude is encountered.	Change cyclically as a variable of underlying environmental and internal influences; any drastic changes are smoothed across previous mood states	Intensity changes in short-term peaks as eliciting stimuli appear, disappear, and change distance; habituation effects describe decay of emotion even in the presence of stimuli.
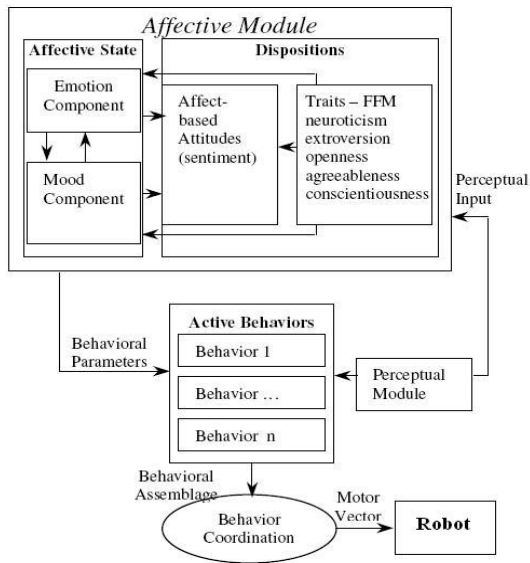
Fig. 1. Conceptual View of TAME

3.2 Psychological and Mathematical Foundations

The mathematical foundations for TAME have been significantly revised since its first publication [10]; detailed information on activation and change of individual components can be found in [11], while the focus of the following section will be on the effect on behavior and its applicability to humanoid robots.

3.2.1 Personality Traits

Personality defines an organism's recurrent patterns of behavior and emotionality. The Big Five taxonomy of personality traits [12] was chosen for this component for its universality: it is consistent over time, cultures, ages, and even applicable to non-human animals. The taxonomy has 5 broad dimensions, each of which is further sub-divided into facets; therefore robot's personality can be as simple or as complex as desired. Traits provide a two-fold advantage for humanoid robots: first, they serve a predictive purpose, allowing humans to understand and infer robot's behavior better; second, they allow adaptation to different tasks and environments, where certain trait configurations are better suited for one or another task or environment.

The five global dimensions are *Openness, Conscientiousness, Extraversion*, *Agreeableness* and *Neuroticism*; each of them has its own effect on robot behavior. For example, in a humanoid, extraversion could be expressed by keeping a closer distance to the human, frequent "smiles", more gestures, etc.; this trait would be appropriate for tasks requiring engagement and entertainment from a robot, e.g., a museum guide or a play partner for kids. The traits are modeled as vectors of intensity, where intensity refers to the extent to which a trait is represented. In the robot, these intensities: are defined a priori by a human; don't change throughout the robot's "life" (this could be a single run, an interaction with a person, or the robot's entire physical life-span); and are not influenced by any affective phenomena. We provide a functional mapping from the trait space onto behavioral parameter space as a 2nd degree polynomial, where 3 pairs of corresponding data points are minimum trait/parameter, maximum, and default/average (the values are taken from the normally distributed human psychological data [13]).

Traits can have a direct or an inverse influence on particular behaviors, and this relationship is defined in a matrix beforehand. In cases where multiple traits affect the same behavior (e.g., Neuroticism may push the robot away from the obstacles while Conscientiousness make it go closer for a faster route), first a trait/parameter mapping is calculated, according to the chosen function $f_{ij}(p_j)$, where trait i influences behavior j, a polynomial in this case. Then, the results are averaged across all influencing personality traits to produce the final parameter value used thereafter:

$$B_j = \frac{1}{\sum\limits_{i=1}^{N} \left| pb_{ij} \right|} \sum\limits_{i=1}^{N} f_{ij}(p_i) \qquad (1)$$

where B_j is a particular behavioral parameter, $f_{ij}(p_i)$ is the function that maps personality trait p_i to B_j, N is the total number of traits, and $\vec{pb}$ is personality/behavior dependency matrix; if there is no influence, the result of $f_{ij} = 0$.

3.2.2 Emotions

From an evolutionary point of view, emotions provide a fast, flexible, adaptive response to environmental contingencies. They appear as short-term, high-intensity peaks in response to relevant stimuli (we don't usually live in a constant flux of emotions), and serve a number of functions, of which most applicable for humanoids are communicative, expressive and affiliative, e.g., fear communicates danger and a request for help, while joy in response to a bright smile helps forge trust and camaraderie. The primary,

reactive emotions of *fear, anger, disgust, sadness, joy and interest* were chosen, in part because these basic emotions have universal, well-defined facial expressions, are straightforwardly elicited, and would be expected, perhaps subconsciously, on a humanoid's face, as appearance does affect expectations.

From an emotion generation point of view, Picard [14] suggests a number of properties desirable in an affective system: 1) a property of activation – emotion is not generated below a certain stimulus strength; 2) a property of saturation – refers to an upper bound; 3) a property of response decay; and 4) a property of linearity – emotion generation will approximate linearity for certain stimulus strength range. Taking these properties into consideration, the resulting function for emotion generation (based on stimulus strength) resembles a sigmoid, in which the left side corresponds to activation, the right side corresponds to saturation (amplitude), and the middle models the actual response.

Emotions are also highly dependent on traits and moods: personality may influence the threshold of eliciting stimulus (activation point), peak (amplitude) and rise time to peak (affecting the slope of the generation curve); and moods can vary the threshold of experiencing a specific emotion. For example, Extraversion is correlated with positive emotions, therefore a humanoid robot high in this dimension would display more smiles, excited gestures and other expressions of joy. Attitude also has an effect on emotion – the object of like or dislike may serve as a stimulus for emotion generation.

Emotions can have a varied impact on behavior, from a subtle slowing to avoid a disgustful object to a drastic flight in response to extreme fear. This effect can be modeled by linear mapping from emotion strength to relevant behavioral parameters, and Fig. 2 provides a comparative view across time of stimulus strength (an object appears, comes closer, and then is gone), corresponding emotion activation (after response decay and smoothing), and the Object Avoidance Gain (which causes an avoidance response to Fear); duration is plotted along the x axis, and normalized values for stimulus strength, fear and object avoidance gain along the y axis.
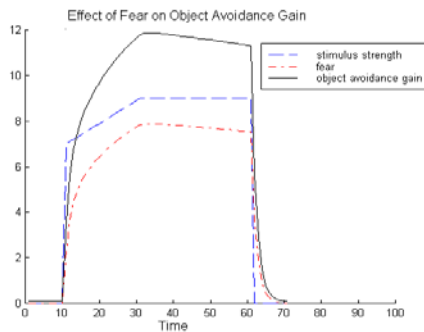

Fig. 2. Example of Fear to Object Avoidance Gain Mapping

3.2.3 Moods

Unlike emotions, moods represent a global, continuous affective state, cyclically changing and subtle in expression. Mood can be represented along two dimensions, Positive Affect and Negative Affect [15], where Negative Affect refers to the extent to which an individual is presently upset or distressed, and Positive Affect generally

refers to one's current level of pleasure and enthusiasm. The level of arousal for both categories can vary from low to high; a low positive mood value has a negative connotation ("sluggish", "disinterested") and refers to insufficient level of pleasure and enthusiasm, rather than just low.

In humanoids, cyclical variations in moods over time can be determined based on the underlying variations in environmental and internal conditions (such as light and battery levels) with any sudden changes smoothed out by taking into consideration a number of prior mood states – filtering over a longer period of time results in slower and smaller mood changes and helps tone down any drastic spikes due to emotions.

Moods are mild by definition, and would only produce a mild, incremental effect, or a slight bias, on the currently active behaviors. Moods can have a direct or inverse influence on a behavioral parameter. A behavior-mood dependency matrix $\overrightarrow{mb} = [mb_{ij}]$, where $mb_{ij} \in \{-1,0,1\}$ is defined, where –1 corresponds to inverse influence, +1 to direct influence, and 0 to absence of mood influence on behavior. Positive and negative mood may influence the same behavioral parameters, and this influence is treated as additive. As moods are updated continuously, new mood-based values of behavioral gains/parameters replace the existing trait-based values in the following manner:

$$B_{i,mood} = B_{i,trait} + K_i \sum_{j=1}^{N} mb_{ji} \cdot m_j \qquad (1)$$

where $B_{i,mood}$ is the updated behavioral parameter i, mb_{ij} is the mood-behavior dependency matrix value for mood j, m_j is the current value of mood j, N is the total number of mood categories (2), and K is a scaling factor to ensure that the moods produce only incremental effect as opposed to overpowering any of the parameters.

Fig. 3 shows an example of incremental effects of moods on behavior. Let's suppose mood can bias robot's obstacle avoidance behavior. For example, if visibility is poor, it may be advantageous to stay farther away from obstacles to accommodate sensor error, and vice versa, in good visibility it may be better to concentrate on task
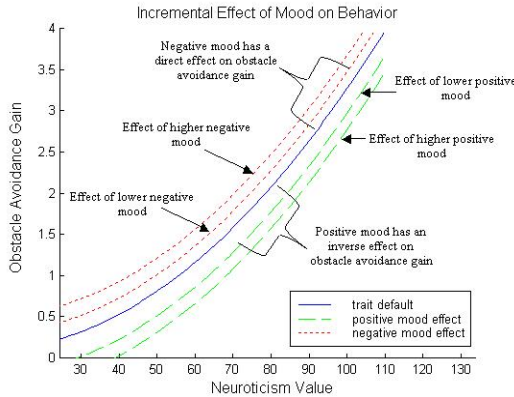

Fig. 3. Direct/Inverse Mood Effects on Behavior at Different Neuroticism Values

performance. Thus, negative mood can bias the obstacle avoidance gain by raising it, and positive mood by lowering it. Neuroticism also affects it by setting the default parameters to be used throughout the life-cycle, and the incremental effect of moods is shown against the space of trait-based defaults (plotted in solid blue center line).

3.2.4 Affective Attitudes

From a multitude of definitions of attitudes, the following was adopted as the working definition for TAME: "a general and enduring positive or negative feeling about some person, object or issue" [16]. It stresses relative time-invariance ("enduring"), object/situation specificity, and the role of affect/affective evaluation in the attitude concept. Attitude can be represented on a single axis, with 0 being a neutral attitude, negative values referring to an increasingly strong negative attitude (ranging from a mild dislike to hatred), and positive values to an increasingly strong positive attitude (e.g., from a subtle like to adoration). Affective attitudes are closely related to emotions, and may even originate in one; therefore initially they will be expressed through a related emotion, and not a behavioral change per se, determining both the type of emotion invoked and its intensity. For humanoids, attitudes would be invaluable in establishing rapport and understanding with interacting humans, as the robot would respond to people's behavior towards it by, for example, expressing joy at the sight of someone it "likes", or sadness when they leave.

4 Architectural Design and Implementation

For our initial research we will test the TAME model within the Georgia Tech *MissionLab*[1] [17] software system and prototype on a Nao humanoid robot (Fig. 4 left). The architectural design overview appears in Figure 4 Right.

There are three major tasks to be completed from a software perspective: (1) Integration of *MissionLab*'s Hserver and the API for Nao; (2) Development of a stand-alone version of TAME running as a separate process within *MissionLab*, capable of updating the externalized affective variables based on a range of conditions including time passage and external stimuli; and (3) The creation of an appropriate set of humanoid behaviors. Suitable perceptual algorithms for the Nao robot will need to be developed as well as connections to the externalized motivational variables already resident in *MissionLab*. The generic functional description for each of these behaviors is:

$$\beta \, (\mathbf{s},\mathbf{a}) \Rightarrow \mathbf{r}$$

where β is the behavioral function mapping the stimulus $\mathbf{s}$ and affective value $\mathbf{a}$ onto an instantaneous response $\mathbf{r}$.

Specific behavior selection, specification, and implementation will be driven from the demonstration scenarios, while the affective models will be derived from both cognitive science models and then through empirical observational studies of the behavior of the humanoid. As is standard for *MissionLab*, these behaviors will be encoded using CDL (Configuration Description Language) and CNL (Configuration Network Language) [17]. The behavioral specifications for the scenarios will be

---

[1] MissionLab is freely available for research and education at:
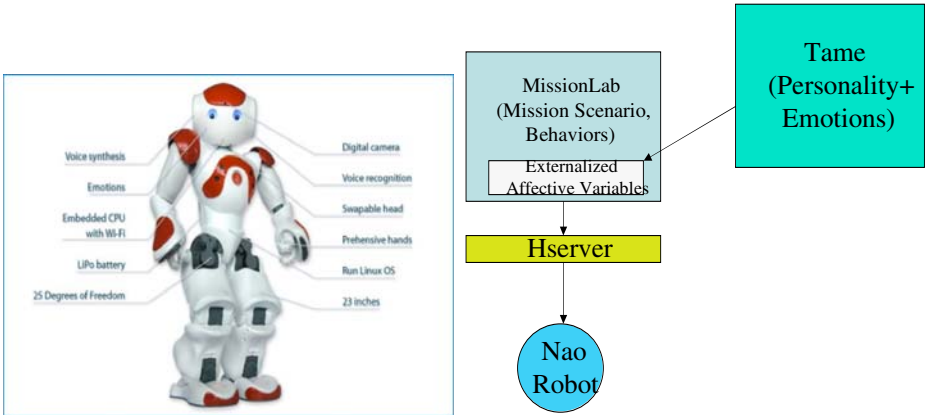http://www.cc.gatech.edu/ai/robot-lab/research/MissionLab/

Fig. 4. (Left) Nao Robot (source Aldebaran Robotics) (Right) TAME integrated with *MissionLab*

represented as finite state acceptors (FSAs) within the configuration editor (cfgedit) component of *MissionLab*. Demonstrations will be created to illustrate the effectiveness of TAME by conveying the appearance of various forms of affect to a human observer, to be developed into full human-robot interaction (HRI) studies.

References

[1] Arkin, R.C.: Moving Up the Food Chain: Motivation and Emotion in Behavior-based Robots. In: Fellous, J., Arbib, M. (eds.) Who Needs Emotions: The Brain Meets the Robot, Oxford University Press, Oxford (2005)

[2] Brooks, A., Arkin, R.C.: Behavioral Overlays for Non-Verbal Communication Expression on a Humanoid Robot. Autonomous Robots 22(1), 55–75 (2007)

[3] Arkin, R., Fujita, M., Takagi, T., Hasegawa, R.: An Ethological and Emotional Basis for Human-Robot Interaction. Robotics and Autonomous Systems 42(3-4) (March 2003)

[4] Miwa, H., Takanishi, A., Takanobu, H.: Experimental Study on Robot Personality for Humanoid Head Robot. In: IEEE IROS, pp. 1183–1188 (2001)

[5] Fukuda, T., Jung, M., Nakashima, M., Arai, F., Hasegawa, Y.: Facial expressive robotic head system for human-robot communication and its application in home environment. Proc. of the IEEE 92, 1851–1865 (2004)

[6] Breazeal, C.: Emotion and Sociable Humanoid Robots. International Journal of Human-Computer Studies 59, 119–155 (2003)

[7] Esau, N., Kleinjohann, B., Kleinjohann, L., Stichling, D.: MEXI: machine with emotionally eXtended intelligence. Design and application of hybrid intelligent systems, 961–970 (2003)

[8] Moshkina, L., Arkin, R.C.: Human Perspective on Affective Robotic Behavior: A Longitudinal Study. In: IEEE IROS (2005)

[9] Arkin, R.C.: Behavior-based Robotics. MIT Press, Cambridge (1998)

[10] Moshkina, L., Arkin, R.C.: On TAMEing Robots. In: Interanational Conference on Systems, Man and Cybernetics (2003)

[11] Moshkina, L.: An Integrative Framework for Affective Agent Behavior. In: IASTED International Conference on Intelligent Systems and Control (2006)
[12] McCrae, R.R., Costa, P.T.: Toward a new generation of personality theories: theoretical contexts for the five-factor model. In: Five-Factor Model of Personality, pp. 51–87 (1996)
[13] Costa, P.T., McCrae, R.R.: NEO PI-R Professional Manual. Psychological Assessment Resources (1992)
[14] Picard, R.W.: Affective Computing. MIT Press, Cambridge (1997)
[15] Watson, D., Clark, L.A., Tellegen, A.: Mood and Temperament. The Guilford Press (2000)
[16] Breckler, S.J., Wiggins, E.C.: On Defining Attitude and Attitude Theory: Once More with Feeling. In: Pratkanis, A.R., Breckler, S.J., Greenwald, A.G. (eds.) Attitude Structure and Function, pp. 407–429. Lawrence Erlbaum Associates, Mahwah (1989)
[17] MacKenzie, D., Arkin, R.C., Cameron, J.: Multiagent Mission Specification and Execution. Autonomous Robots 4(1), 29–57 (1997)