An Open Architecture for Emotion and Behavior Control of Autonomous Agents

Juan D. Velásquez
MIT Artificial Intelligence Laboratory
545 Technology Square,
Room NE43-935
Cambridge, MA 02139
jvelas@ai.mit.edu

Masahiro Fujita
D21 Laboratory,
Sony Corporation
6-7-35 Kitashinagawa,
Shinagawa-ku, Tokyo 141 Japan
mfujita@pdp.crl.sony.co.jp

Hiroaki Kitano
Sony Computer Science
Laboratory, Inc.
Takanawa Muse Building,
3-14-13 Higashi-gotanda
Shinagawa-ku, Tokyo, 141 Japan
kitano@csl.sony.co.jp

1. ABSTRACT

We present an open architecture for autonomous agents that combines emotions, drives, and action selection. We show how emotions constitute the agent's main motivational system and influence how behaviors are selected and controlled. The architecture has been designed and implemented as part of an object-oriented framework, portable among different agent platforms. These ideas have been used to develop and control different types of autonomous agents, including synthetic agents and physical robots, that exhibit emotional behaviors.

2. INTRODUCTION

Although some models and agent architectures that consider emotion have been proposed [1, 2, 4, 5, 6], in general, the consideration of emotions from a computational perspective, and in particular in the field of Artificial Intelligence, has been largely ignored.

In this paper, we propose a general, open architecture that integrates emotions, drives, and behaviors for the control of autonomous agents. By open, we mean that the architecture has been designed to be portable in the sense that it can be used to develop and control different types of agents, whether they are synthetic agents or physical robots. And by general, we mean that the underlying model of emotions and action selection can be used not only to model emotional agents for entertainment purposes, but also to develop and test new and old theories of emotions and intelligence.

3. THE ARCHITECTURE

The architecture described here is derived from previous research on computational models of emotion [7], but it incorporates several important additions and improvements, especially with respect to the Behavior System and the algorithm for behavior selection and control.

3.1 Drives

Drives represent needs that motivate the agent into action. Each Drive includes a set of Releasers which obtain sensory data and identify special conditions which will either increase or decrease the value of the entity they belong to. In the case of drives, these releasers represent control systems which regulate a controlled variable that is maintained within a certain range. If the value for this variable does not match a specified set point, an error signal is produced. This error signal is fed to the appropriate drive, in which it can be combined with error signals from other relevant control systems.

3.2 Emotions

The Emotion Generation System consists of a distributed network of self-interested emotional systems representing different emotion families, including *Anger, Fear, Distress/ Sadness, Enjoyment/Happiness, Disgust*, and *Surprise*. For details on why this set of emotions was chosen and not other, the reader may consult [7].

Like Drives, Emotional systems also have special kinds of Releasers which are constantly monitoring the environment to check for the appropriate conditions that would elicit the emotion they represent. Each emotional system receives input from these releasers, as well as from other emotional systems. Since both drives and emotional systems are implemented using the same abstraction, it is easy for emotions to interact with drives and vice-versa. For instance, high levels of *Hunger* might increase the level of *Distress* and *Anger*. Similarly, high levels of *Sadness* might decrease *Hunger*.

3.3 Behaviors

The Behavior System is a distributed network of self-interested behaviors which can be organized into hierarchies of groups. Each behavior is responsible for determining its relevance given the agent's motivational state (i.e. the state of its drives and emotional systems) and any relevant external stimuli from the world. The value of each behavior is computed based on the following update rule:

$$B_{jt} = \delta(B_{jt-1}) + \alpha(B_{jt-1}) + \frac{1}{N} \cdot \sum_{n}(R_{nj} \cdot W_{nj}) + \sum_{l}(G_{lj} \cdot B_{lt}) - \sum_{m}(H_{mj} \cdot B_{mt}) \tag{1}$$

Where B_{jt} is the value of Behavior j at time t; B_{jt-1} is the value of Behavior j at the previous time step; α is an auto-excitation function (used to model persistence) associated with Behavior j; δ is a decay function (used to model disinterest) associated with Behavior j; R_{nj} is the value of releaser n and W_{nj} is the weight for releaser n, where n ranges over the N releasers for Behavior j; G_{lj} is the Excitatory Gain that Behavior l applies to Behavior j, and B_{lt} is the intensity of Behavior l at time t, where l ranges over the set of behaviors that excite Behavior j; H_{mj} is the Inhibitory Gain that Behavior m applies to Behavior j, and B_{mt} is the intensity of Behavior m at time t, where m ranges over the set of behaviors that inhibit Behavior j.

4. SAMPLE APPLICATIONS

The architecture described above has been implemented as an object-oriented framework for the development of emotional agents. This framework has been designed so that it can be used to control different sorts of agents, including both synthetic agents, such as Simón the Toddler (shown in Figure 1.), and physical robots, such as MUTANT [3], a quadruped pet-type entertainment robot (shown in Figure 2.).

Besides the pre-defined set of drives and emotions, every agent has a repertoire of behaviors, such as "*sleep*", "*play with toy*", "*wag the tail*", "*eat*", "*avoid obstacles*", and so on, which correspond to action tendencies in response to the state of its motivational systems.

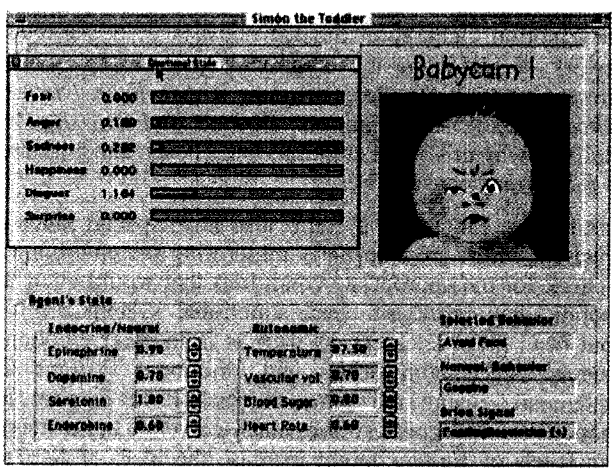

Figure 1. Simón the Toddler: An Animated Synthetic Agent

Users interact with the agents in different ways. In the case of animated agents, the user performs actions in its environment using sliders, icons, buttons, and menus. Similarly, to interact with its users and exhibit emotional behavior, MUTANT has been equipped with a Micro-Camera-Unit (MCU), two microphones that act as audio sensors, tactile sensors in the head and tail, and a loudspeaker.

Figure 2. MUTANT: A Pet-Type Robot

5. CONCLUSIONS

We have presented an open architecture that integrates emotions, drives, and behaviors for the control of autonomous agents. A more robust Behavior System has been integrated into the model of emotions previously proposed by [7]. The resulting system is a complete architecture that models and integrates emotions with other motivational states to influence how behaviors are selected and controlled.

6. REFERENCES

[1] Blumberg, B. 1994. Action-Selection in Hamsterdam: Lessons from Ethology. In *Proceedings of SAB 94*, 108-117. Brighton, England: MIT Press.

[2] Elliot, C.D. 1992. The Affective Reasoner. Ph.D. Thesis, Institute for the Learning Sciences, Northwestern University.

[3] Fujita, M., and Kageyama, K. 1997. An Open Architecture for Robot Entertainment. In: *Proceedings of Agents 97*. Marina del Rey, CA: ACM Press.

[4] Kitano, H. 1995. A Model for Hormonal Modulation of Learning. In: *Proceedings of IJCAI-95*. Montreal.

[5] Pfeifer, R. 1988. Artificial Intelligence Models of Emotion. In: Hamilton, V., Bower, G. H., and Frijda, N. eds. *Cognitive Perspectives on Emotion and Motivation*, 287-320. Netherlands: Kluwer.

[6] Reilly S. 1996. Believable Social and Emotional Agents. Technical Report, CMU-CS-96-138, School of Computer Science, Carnegie Mellon University.

[7] Velásquez, J. 1997. Modeling Emotions and Other Motivations in Synthetic Agents. In: *Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI-97)*. Providence, RI: AAAI/MIT Press.