Proceedings ACII 2005, 1st International Conference on Affective Computing and Intelligent Interaction, October 22-24, 2005, Beijing, China

A Three-Layered Architecture for Socially Intelligent Agents: Modeling the Multilevel Process of Emotions

Christine L. Lisetti[1],
Andreas Marpaung[2]

[1] Affective Social Computing Laboratory, Institut Eurecom, , Postfach 10 52 80,
06560 Sophia-Antipolis, France
{Lisetti}@eurecom.fr
http://www.eurecom.fr/~lisetti
[2] Computer Science Dept., University of Central Florida,
Orlando, USA
{marpaung}@cs.ucf.edu

Abstract. In this article, we propose the design of a three-layered agent architecture inspired from the Multilevel Process Theory of Emotion (Leventhal and Scherer, 1987). Our project aims at modeling emotions on an autonomous embodied robotic agent, expanding upon our previous work (Lisetti, et al., 2004). Our agent is designed to socially interact with humans, navigating in an office suite environment, and engaging people in social interactions. We describe: (1) the psychological theory of emotion which inspired our design, (2) our proposed agent architecture, (3) the needed hardware additions that we implemented on a robot, (3) the robot's multi-modal interface designed especially to engage humans in natural (and hopefully pleasant) social interactions.

1 Introduction

As robots begin to enter our everyday life, an important human-robot interaction issue becomes that of *social interactions*. Because emotions have a crucial evolutionary functional aspect in social intelligence, without which complex intelligent systems with limited resources cannot function efficiently [13], [14] or maintain a satisfactory relationship with their environment [15], we focus our current contribution to the study of emotional social intelligence for robots. Indeed, the recent emergence of affective computing combined with artificial intelligence [16] has made it possible to design computer systems that have "social expertise" in order to be more autonomous and to naturally bring the human – a principally social animal – into the loop of human-computer interaction.

In this article, *social expertise* is considered in terms of (1) internal motivational goal-based abilities and (2) external communicative behavior. Because of the important functional role that emotions play in human decision-making and in human-human communication, we propose a paradigm for modeling some of the functions of emotions in intelligent autonomous artificial agents to enhance both (a) robot autonomy and (b) human-robot interaction. To this end, we developed an autonomous service robot whose functionality has been designed so that it could socially interact with humans on a daily basis in the context of an office suite environment and studied and evaluated the design *in vivo*. The social robot is furthermore evaluated from a social informatics approach, using workplace ethnography to guide its design *while* it is being developed.

2 A Three-Layered Emotion-Based Architecture

2.1 Related Research

There have been several attempts to model emotions in software agents and robots and to use these models to enhance functionality. El-Nasr [17] uses a fuzzy logic model for simulating emotional behaviors in an animated environment. Contrary to our approach directed toward robots, her research is directed toward HCI and computer simulation.

Velasquez's work [10], [18] is concerned with autonomous agents, particularly robots in which control arises from emotional processing. This work describes an emotion-based control framework and focuses on affect programs which are implemented by integration of circuits from several systems that mediate perception, attention, motivation, emotion, behavior, and motor control. These range from simple reflex-like emotions, to facilitation of attention, to emotional learning. Although the approach is different, its motivation is similar to ours.

Breazeal's work [8], [9] also involves robot architectures with a motivational system that associates motivations with both drives and emotions. Emotions are implemented in a framework very similar to that of Velasquez's work but Breazeal's emphasis is on the function of emotions in social exchanges and learning with a human caretaker. Our approach is different from Breazeal's in that it is currently focused on both social exchanges and the use of emotions to control a single agent.

In Michaud's work [19], [20], emotions *per se* are not represented in the model, but emotion capability is achieved by incorporating it into the control architecture as a global background state. Our approach which chooses to represent the emotional system explicitly (as discussed later) differs from Michaud's in that respect. Although both Michaud and our approach revolve around the notion of emotion as monitoring progress toward goals, our work explicitly represents emotion and corresponds to a formal cognitive model.

Murphy and Lisetti's approach [11] uses the multilevel hierarchy of emotions where emotions both modify active behaviors at the sensory-motor level and change the set of active behaviors at the schematic level for a pair of cooperating heterogeneous robots with interdependent tasks.

Our current approach builds on that work, setting the framework for more elaborate emotion representations while starting to implement simple ones and associating these with expressions (facial and spoken) in order to simultaneously evaluate human perceptions of such social robots so as to guide further design decisions.

2.2 A Multilevel Emotion State Generator (ESG)

The design of our emotion-based architecture is based on the *Multilevel Process Theory of Emotion* (Leventhal and Scherer, 1987) shown in Figure 1. In short, the Multilevel Process Theory of Emotion postulates that the experience of emotion is a product of a hierarchical multi-component processing system.: (1) *Sensory motor level* – generates the primary emotion in response to the basic stimulus features in a non-deliberative manner; (2) *Schematic level* – integrates specific situational perceptions with autonomic, subjective, expressive and instrumental responses in a concrete and patterned script-like memory system; (3) C*onceptual level* – is deliberative and involves reasoning over the past, projecting into the future, and comparing emotional schemata to avoid unsuccessful situations (and is not be treated in our current implementation).

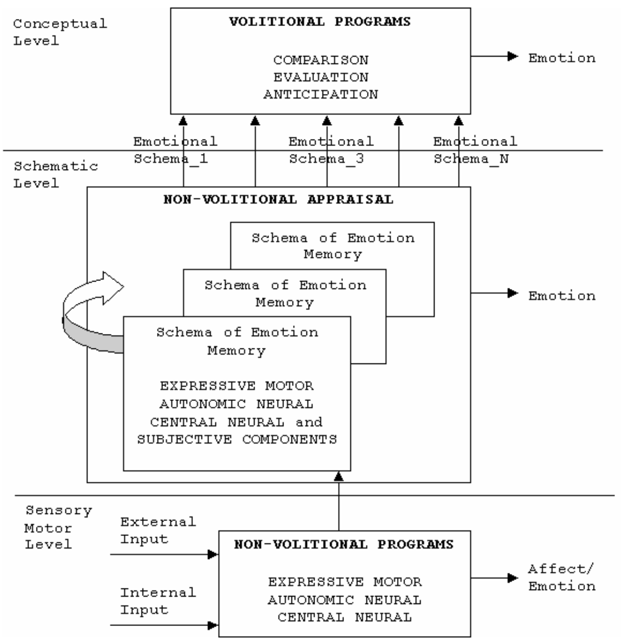

Figure 1: Multilevel Emotion State Generator (ESG)

2.2 Robotic Platform

Our robotic implementation, a Pioneer PeopleBot from Activmedia shown in Figure 2, is a continuation of our previous project – Cherry the AmigoBot – the little red robot designed to navigate in an office suite, to deliver supplies to offices and to greet people according to their social status via an animated avatar (Lisetti et al., 2004).

Proceedings ACII 2005, 1st International Conference on Affective Computing and Intelligent Interaction, October 22-24, 2005, Beijing, China

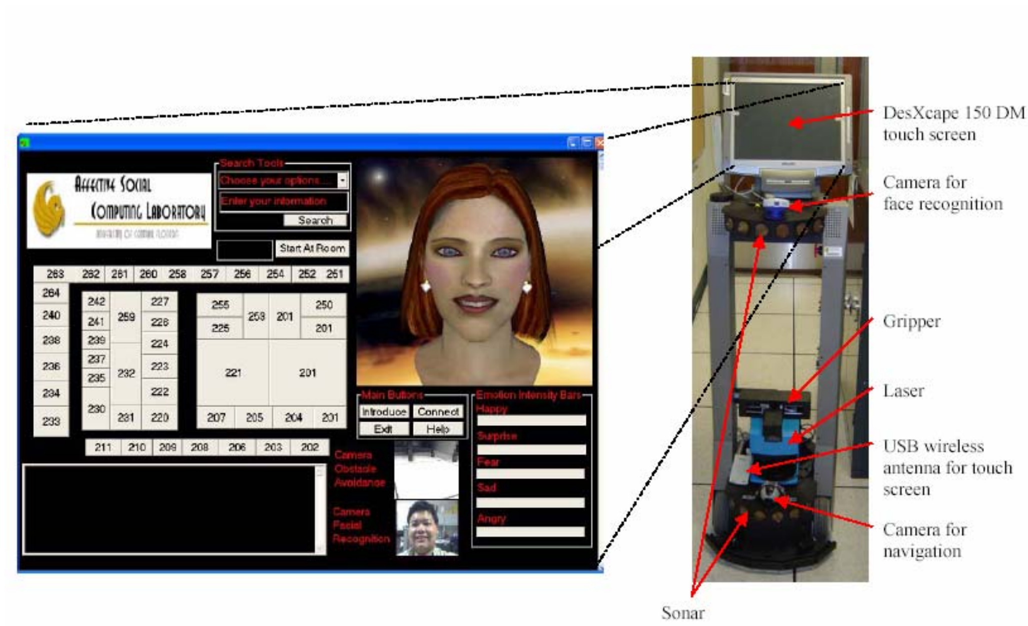

Figure 2: Our social robot platform

The ActivMedia Peoplebots (ActivMedia, 2002) are equipped with twenty-four sonars, one movable camera, one gripper that can be moved up and down, front and back. To meet our goals of enabling human-centered dynamic interactions with our agent, we added hardware to the original PeopleBot: a DesXcape 150 DM touch-screen so that users can easily point-and-click the destination to the robot assistant, a wireless antenna for DesXcape touch screen communication, one floor-level camera for navigation, one eye-level camera for face-recognition, (and one laser for future ability to generate maps instantly).

2.3 Multimodal interface with an Expressive Avatar

Through the on-board computer, we execute the interface that we created and display it through the touch screen wirelessly (Figure 2). The interface, a modified version of Cherry's which was tested as far as its users' acceptance (Lisetti et al., 2004), integrates several components such as an anthropomorphic avatar, a point-and-click map, progress bars showing real-time changes of the robot's emotion-like states, in addition to several help menus, i.e., speech text box, search properties, start-at-room option, and two live-capture video frames – one showing the robot's floor vision, the other its eye-level vision.

In addition, in order to enable our agent to have the ability to communicate its internal emotion-like states, we animate an avatar 's facial expressions. For each emotion-like stage we are modeling (happiness, surprise, fear, sad, and anger), we have designed facial expressions closed to the Facial Action Coding System (FACS) which are shown in Figure 3 (a-c).

(a) (b) (c)

Figure 3: Facial expressions for the modeled emotions (a) Happy, (b) Sad, (c) Angry

2.3 Using the Architecture for our Robotic Platform

The general agent architecture is shown in Figure 4. For every cycle (in our case, it is 1000 mm travel distance), the Input Sensors send inputs read to the Perceptual System for processing. The Perceptual System is a simple system that abstracts information from the outside world with some interpreted meaning for the robot (e.g., walls, floors, open or closed doors, faces) and sends its outputs (valid sonar readings, vision-navigation interpretation, and person's name) to the Emotion State Generator (ESG) which in turn triggers certain emotion-like states as we will describe later.

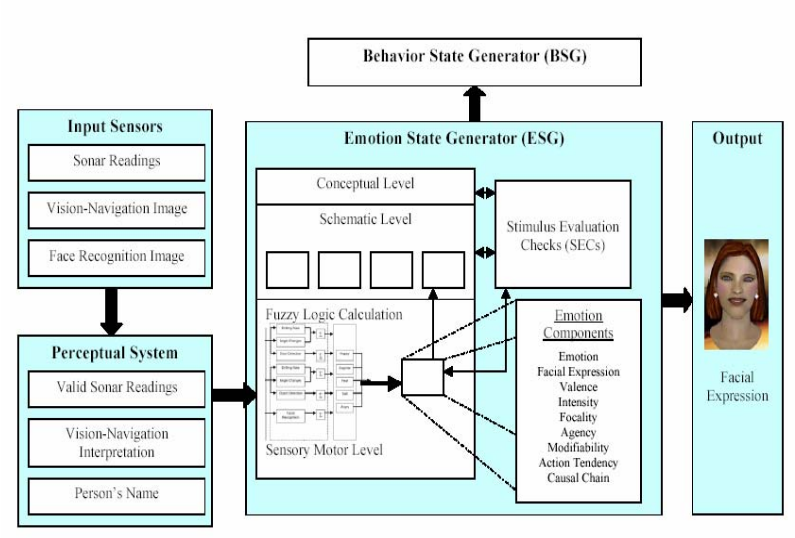

Figure 4: Petra's Detailed Three-Layered Architecture

In our design, sonar is used in conjunction with vision to navigate around the office suite. A sonar reading is considered invalid if the sum of the left-most and the right-most sonar readings are extremely more or less than the distance between the aisle (1,500 mm for our case). And vice versa, the reading is valid if the sum of both readings is around 1,500 mm. Floor vision is used to perform course correction to center itself between the corridor aisles, and to detect obstacles (e.g. garbage can, boxes, people). Eye-level vision performs face recognition on a database of the facial images of the 25 workers in the office suite.

As shown in Figure 4, the Perceptual System passes these sensory information to the Emotional State Generator (ESG): door detected, object detected, whether a person was recognized, the name of the person recognized. A fuzzy logic model (Takagi, Sugeno, 1985) is further used to deal with the uncertainty of the sensory information and determine the current emotion-like state based on an (currently) ad-hoc OR-mapping.

For example, happiness is strongly correlated with low variations from the sonar reading, recognizing a face, or finding that the office where it was sent is open. Similarly, anger is correlated with finding repetitively closed doors and repetitively not recognizing people.

2.3 Sequential Evaluation Checks

In order to produce emotion for each level, many researchers have hypothesized that specific emotions are triggered through a series of sequential evaluation checks (SECs) of the various stimuli (Scherer, 1984; Scherer, 1986; Weiner, Russell, and Lerman, 1979; Smith and Ellsworth, 1985). We have identified specific SECs we considered particularly useful for the design of artificial agents into an Affective Knowledge of Representation (AKR) (Lisetti, 2002). In AKR, each emotion has many components, e.g., valence, intensity, focality, agency, modifiability, action tendency, and causal chains which we describe below:

Facial Expression: *happy/sad/surprised/angry:* is used to control the avatar's animation.
Valence: *positive/ negative:* is used to describe the pleasant or unpleasant dimension of an affective state.

Intensity: *very high/ high/ medium/ low/ very low:* varies in terms of degree. The intensity of an affective state is relevant to the importance, relevance and urgency of the message that the state carries.

Focality: *event/ object:* is used to indicate whether the emotions are about something: an event (the trigger to surprise) or an object (the object of jealousy).

Agency: *self/ other:* is used to indicate who was responsible for the emotion, the agent itself *self*, or someone else *other*.

Modifiability: *high/ medium/ low/ none:* is used to refer to duration and time perspective, or to the judgment that a course of events is capable of changing.

Action tendency: identifies the most appropriate (suite of) actions to be taken from that emotional state. For example, happy is associated with generalized readiness, frustration with change current strategy, and discouraged with give up or release expectations.

Causal chain: identifies the causation of a stimulus event associated with the emotion. For example, happy has these causal chains: (1) Something good happened to me, (2) I wanted this, (3) I do not want other things, and (4) Because of this, I feel

good. These can used by the agent to verbally express and explain to humans what emotion-like state it finds itself in.
Emotion components are checked to create a schema of emotion to be stored in the memory. The checking is done by assigning appropriate values to the emotion components based on pleasantness, importance, relevance, urgency, etc.

Table 2 shows a schema when an unexpected moving object suddenly appears in the captured navigation-image, i.e, walking students. In this case, surprise will become the current state. Indeed, the sudden appearance of a person in the navigation image is detected as an obstacle that can slow down the navigation process, due to the course correction that needs to be performed should the person remain in the navigation image on the next cycle. Thus *intensity* component is very high and the *action tendency* is to avoid potential obstacles. Since the face cannot be detected at farther distance, the *valence* is negative, and the *modifiability* is set to its default–medium because the agent has not performed obstacle avoidance to change the course event. In addition, the facial expression component is set to surprise to drive the animation of the facial expressions of the interface avatar.

Table 2: *Surprise* schema

Components	**Values**
Emotion	SURPRISE
Facial Expression	Surprise
Valence	Negative
Intensity	Very High
Focality	Object – walking student
Agency	Other
Modifiability	Medium
Action Tendency	Avoid
Causal Chain	- Something happened now - I did not think before now that this will happen - If I thought about it, I would have said that this will not happen - Because of this, I feel something bad

3 Conclusion

This work has presented an initial attempt to develop a 3-layered emotion-based architecture for artificial agents based on Scherer and Leventhal's multilevel process theory of emotions. Still much work remains to render this approach complete, and this study will continue to point out the power of emotions in human intelligence (which might be impossible to fully computationalize).

References

ActivMedia. www.activmedia.com, 2002.

Arkin, R. C. *Behavior-Based Robotics*, Cambridge, MA: MIT Press, 1998.

Breazeal, C. and Scassellati, B., "Infant-like Social Interactions Between a Robot and a Human Caretaker". To appear in *Special issue of Adaptive Behavior on Simulation Models of Social Agents*, guest editor Kerstin Dautenhahn. 2000

Breazeal, C. "Emotion and sociable humanoid robots". *International Journal of Human Computer Studies.* Vol. 59. pg. 119 – 155. 2003

Brooks, R and Flynn, A. "Fast, Cheap, and Out of Control", AI Memo 1182, MIT AI Laboratory, 1989.

Brown, S. Lisetti, C. and Marpaung, A. Cherry, the Little Red Robot with a Mission and a Personality. In *Working Notes of the AAAI Fall Symposium Series on Human-Robot Interaction*, Menlo Park, CA: AAAI Press. Cape Cod, MA, November 2002.

Casper, J. "Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center." MS Thesis, Computer Science and Engineering, University of South Florida, April 2002.

Casper, J. and Murphy, R. "Workflow study on human-robot interaction in USAR." In *Proceedings of ICRA 2002,* pp. 1997 – 2003. 2002.

Ekman, P and Friesen, W. "The Facial Action Coding System". Consulting Psychologist Press, San Francisco, CA, 1978.

El-Nasr, Magy Seif. Yen, John. Ioerger, Thomas. FLAME - A Fuzzy Logic Adaptive Model of Emotions, *Automous Agents and Multi-agent Systems,* 3, 219-257, 2000.

Identix Inc. www.identix.com, 2002.

Leventhal, H. “A perceptual-motor processing model of emotion.” In P. Pilner, K. Blankenstein, & I.M. Spigel (Eds.), *Perception of emotion in self and others. Vol. 5* (pp. 1 – 46). New York: Plenum. 1979

Leventhal, H., “Toward a comprehensive theory of emotion.” In L. Berkowitz (Ed.), *Advances in experimental social Psychology. Vol. 13* (pp. 139 – 207). New York: Academic Press, 1980

Leventhal, H. and Scherer, K. “The relationship of emotion to cognition: A functional approach to a semantic controversy.” *Cognition and Emotion.* Vol. 1. No. 1. pp. 3 – 28. 1987.

Lisetti, C. Brown, S. Alvarez, K. and Marpaung, A. A Social Informatics Approach to Human-Robot Interaction with an Office Service Robot. *IEEE Transactions on Systems, Man, and Cybernetics - Special Issue on Human Robot Interaction,* Vol. 34(2), May 2004.

Lisetti, C. and Bianchi, N. Modeling Multimodal Expression of User's Affective Subjective Experience. *User Modeling and User-Adapted Interaction, An International Journal,* Vol. 12(1): 49-84. 2002

Marpaung, A. “Social Robots with Emotion State Generator Enhancing Human-Robot Interaction (HRI)”. Master’s thesis. University of Central Florida. in progress 2004.

Murphy, R. R. Use of Scripts for Coordinating Perception and Action, In *Proceedings of IROS-96* , 1996a.

Murphy, R. R. Biological and Cognitive Foundations of Intelligent Sensor Fusion, *IEEE Transactions on Systems, Man and Cybernetics*, 26 (1), 42-51, 1996b.

Murphy, R. R. Dempster-Shafer Theory for Sensor Fusion in Autonomous Mobile Robots, *IEEE Transactions on Robotics and Automation*, 14 (2), 1998.

Murphy, R. R. *Introduction to AI Robotics*. Cambridge, MA: MIT Press, 2000.

Murphy, R. R., Lisetti, C. L., Irish, L., Tardif, R. and Gage, A., Emotion-Based Control of Cooperating Heterogeneous Mobile Robots, *IEEE Transactions on Robotics and Automation*, Vol. 18, 2002.

Picard, Rosalind W. Affective Computing, Cambridge, Mass.: MIT Press, 1997.

Scherer, K. “Emotion as a multicomponent process: A model and some cross-cultural data.” In P. Shaver (Ed.), *Review of personality and social psychology.* Vol. 5. *Emotions, relationships and health* (pp. 37 – 63). Beverly Hills, CA: Sage. 1984

Scherer, K. “Vocal affect expression: A review and a model for future research.” *Psychological Bulletin.* 99. pp. 143 – 165. 1986

Smith, C. A., Ellsworth, P.C. “Patterns of cognitive appraisal in emotion.” *Journal of Personality and Social Psychology.* 48. pp. 813 – 838. 1985

Simmons, R. Goldberg, D. Goode, A. et al. GRACE An Autonomous Robot for the AAAI Robot Challenge. *AI Magazine.* Vol. 24. No. 2. p. 51 – 72. 2003

Sony. www.us.aibo.com. 2004.

Takagi, H. and Sugeno, M.. "Fuzzy identification of systems and its application to modeling and control". *IEEE Transactions on Systems, Man and Cybernetics*, 15 (1), 1985.

Velasquez, Juan. "Cathexis, A Computational Model for the Generation of Emotions and their Influence in the Behavior of Autonomous Agents". Master's thesis. Massachusetts Institute of Technology. 1996

Velasquez, Juan. "Modeling Emotion-Based Decision Making". *In proceedings of AAAI Fall Symposium Emotional & Intelligent: The Tangled Knot of Cognition.* 1998

Weiner, B., Russell, D., Lerman, D. "The cognition-emotion process in achievement-related contexts. *Journal of Personality and Social Psychology*, 37, pp. 1211 – 1220. 1979.