Beyond Humanoid Emotions: Incorporating Traits, Attitudes and Moods

Lilia Moshkina and Ronald C. Arkin

I. Introduction

No longer does the idea of robot emotions seem far-fetched; not their experiential side, of course, but rather those manifestations of emotion, especially in robots created in human likeness, which would be beneficial for successful interaction with people. Nonetheless, the concept of robot emotions is still a new one, with a myriad of questions to be answered, not the least of which is: What is emotion? In robotics, it is often used as an umbrella term for all things affective, but based on our previous work (see [1] for a summary), we believe that it would be more beneficial to model each affective phenomenon explicitly. Going beyond emotions brings the entire spectrum of affect into play, providing a comprehensive framework with which human-robot interaction could be improved. The robotic framework we propose that combines a number of different phenomena and emphasizes their interconnectedness and synergy is called TAME (Traits, Attitudes, Moods, Emotions). By using TAME, in this paper we'd like to address some of the open questions that arise in the area of implementing and testing humanoid affect.

II. Overview of TAME

The idea behind TAME is simple: in both humans and animals, affect, of which emotions are, though integral, only a part, has been proven beneficial for survival. The same general mechanisms that help us live and prosper may facilitate both effectiveness and acceptance of humanoid robots, if such robots are expected to live among us. In particular, there are four different phenomena that can be classified as affective: personality Traits, affective Attitudes, Moods and Emotions, each performing its own role in humans and having distinct generation mechanisms. One dimension along which they differ is time, including both duration and rate of change. Emotions are the most short-lived of the four, and are fast to rise and fast to decay; moods are longer in duration and change slowly and cyclically; attitudes, once formed, last for a while and are hard to influence; and finally, traits are more or less time-invariant. Another dimension of difference is object-specificity: emotions and attitudes arise in response to a specific object or situation, whereas traits and moods are diffuse, global, and apply at all times. The combination for these four affective types should be especially beneficial for humanoids, as it is suitable for long-term interaction and development of companionship (through, e.g., attitudes).

Psychological and mathematical foundations behind the framework have been discussed in more detail elsewhere ([2, 3, 4]). In brief, the Affective Module containing the aforementioned four components fits within behavior-based robotic control [5] 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 (**Figure 1**).

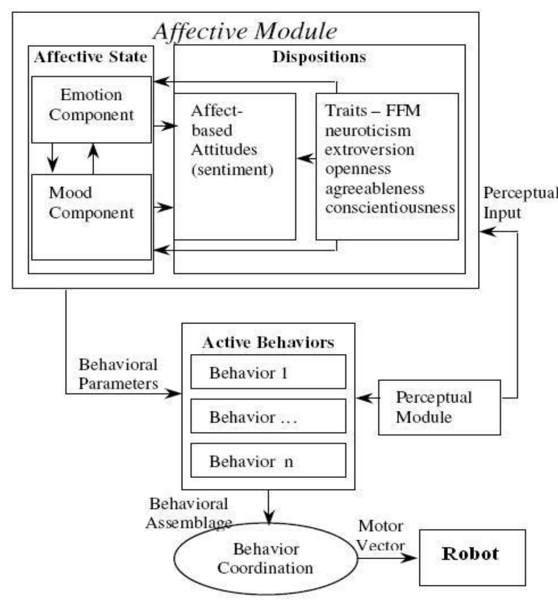

Figure 1: Conceptual View of TAME

III. Challenges of Humanoid Affect

Although TAME covers a wider range of robotic behaviors than application of emotions only, similar implementation and assessment challenges remain. In the rest of the paper we will discuss what we believe these to be, and how they could be successfully addressed.

A. *Subtle and Volatile Nature of Affect*

Unlike other fields of robotics, e.g., vision or gait control, where the goals, tasks and measures are straightforward and

This research is funded under a grant from Samsung Electronics.

The authors are with Georgia Tech Mobile Robot Laboratory, Atlanta, GA, USA 30332; email: {lilia, arkin}@cc.gatech.edu

objective, the advantages of affect are much harder to quantify. In principle, affect influences many spheres of our lives and performs a multitude of functions, but when applied to humanoid robotics, it is often targeted towards communication between humans and robots. What advantages would it bring to human-robot interaction? What kind of tasks would benefit from its inclusion? Are there other functions than communicative that could be useful? What is the best way to show effectiveness of affective components and how to disambiguate between them? Most of these questions are asked in other areas of robotics, but the subtle and volatile nature of affect makes answering them in this case especially challenging.

First of all, emotions are short-term and fleeting, and occur rather infrequently; therefore, it is unlikely that seeing a single expression of, for example, "joy", in the entire interaction would make a large difference to a human. Moods produce only subtle, incremental changes in a robot's behavior as its environment changes, and would not be immediately noticeable to the observer. Attitudes, even though often quite explicit, don't usually form in a single interaction, and traits are best displayed across a variety of tasks and situations. All this: the relative infrequency and short duration of emotions, subtlety of moods, slowly changing nature of attitudes, and constancy of traits make affective phenomena best suited for long-term human-robot interactions. Although this by no means renders short-term robotic affect useless, it nonetheless makes finding appropriate tasks and scenarios non-trivial, especially given that longitudinal studies, though ideal, are very time and resource consuming. Finally, to disambiguate between the advantages of the four affective phenomena present in TAME, an experimental setup akin to lesion studies would be required, where each component is tested separately first, and then in combination; this, again, adds substantially to the complexity of the evaluation process.

B. Recognizing and Comparing Robot Affect

As there are at least two parties to any interaction, before any testing can begin we need to make sure that human interaction participants can correctly "read" any affect exhibited by a robot. Humanoids are not people – so how do we express affective robotic phenomena in a manner that they can be successfully recognized as such by humans? This problem is alleviated in part by our own nature: people treat any interaction partners, including computers, as social actors, applying to them similar social rules, given even minimal cues [7]. Whether or not such cues are sufficient to identify the underlying emotion/mood/trait/attitude can tested in a number of ways. First, a formal user study with a real robot can be conducted with the goal of testing whether the expressions were recognized as intended; this method, though time-consuming, provides the best idea of what worked and what didn't, and how to improve it. Another approach would be to conduct a small pilot study prior to the main one, in which affect recognition would be tested, along with other things; this method would save time and resources, would still involve real robot interaction, but would be more limited in its findings. Finally, an online survey could be done, in which participants would be asked to identify the affective states of a simulated robot, either in an interactive or passive way; this method, though the least costly, would also be the least informative.

A related challenge is comparing robotic affect between various platforms, as different platforms mean different capabilities: what would "fear" look like in a humanoid with a human-like face vs. one without any changeable facial features? Would one platform be easier to work with than the other? How much would the participants be biased by such physical manifestations? For example, physical features of a robot made with an entertainment purpose in mind may provoke an instantaneous affective response, whether or not it was intended by the experimenter [3]. The aforementioned affect recognition testing may in part address this problem, as it would be known ahead of time whether the intended affect was recognized, along with any unintended impressions, therefore these perceived but unintended phenomena can be corrected for later use.

C. Robot Affect Assessment

The challenges described above make determining the benefits of adding affective capabilities to humanoids especially arduous. As such, there are no hard rules or even guidelines for metrics of effectiveness of robotic affect. Most often, it is assessed by purely subjective means or more established psychological or sociological tests; sometimes, observational means are used, and yet more rarely objective measures are employed. The following subsections will discuss these methods and their challenges in more detail.

a) Subjective assessment

Such assessment includes:

- *Self-reported information* in the form of robot- and task-specific questionnaires/interviews, asking the participants about the quality of their interaction, namely how pleasant, easy, and natural it was, and whether they could distinguish any emotions or other affect in the robot. This method allows querying people's perceptions of their interaction, but is very subjective and makes it hard to compare findings from studies by different experimenters.
- *Psychological and sociological measures* (specially developed and validated tests to measure different aspects of interaction). These tests can be used to assess subjects' mood, emotional state, attitudes, presence, acceptance, and many other subjective states. Examples of such measurement scales include: Goldberg's Unipolar Big-Five Markers (personality) [8], Positive/Negative Emotionality Measure (current mood) [9], Self-Assessment Manikin (emotional response) [10], International Affective Picture System

[11], etc. The advantage of these is their documented validity across participants, thus they provide a more or less reliable set of data allowing comparison of users' internal state and perceptions.

b) Observational means

These means can be tentatively divided into purely subjective (qualitative), and those that cross the bridge between identifying an individual's perception of the interaction and distinguishing quantifiable benefits. Provided affect expression in robots is successfully implemented, we would expect people to act differently in response to a humanoid always behaving in a repetitive unemotional way, and in response to one that changes its behavior and expressions according to external and internal stimuli. The list below discusses these methods in more detail:

- *Independent observer assessment* – a person (preferably an ethnographer, sociologist or psychologist) qualitatively characterizes the nature of interaction either real-time or via video. In this case, although user bias is removed, interpretation bias is introduced.
- *Behavioral analysis* – this refers to analysis of micro- and macro-behaviors and speech utterances. In this case, the human-robot interactions are recorded; the behaviors to watch for are carefully selected and accurately described, and then are extracted from the video either automatically, or by independent human coders. For example, suppose that the duration of mutual gaze is a good predictor of the quality of interaction – the longer the mutual gaze episodes, the more pleasant the interaction. Now we have a quantitative measure that would allow us to compare between robots that express affect and those that don't. However, this method still suffers from interpretation bias: the definition of mutual gaze (e.g., angles, acceptable percent of deviation, minimum duration, etc.) needs to be worked out and adapted to the current experiment, and individual differences, such as personality and current state of mind have to be taken into account.
- *Physiological analysis* - certain physiological responses (such as heart rate, skin conductance and temperature) can be measured before, during and after the interaction; such responses can be correlated with subjects' emotional state and arousal level. Though seemingly more objective, this method still suffers from individual differences in responses and low reliability unless the equipment is individually calibrated. Additionally, the equipment is often cumbersome and its presence alone may influence the results.

When combined with self-reported data, these methods can undoubtedly provide a clearer picture of the usefulness of humanoid affect. However, although some researchers believe that such behavioral and physiological measures are objective, it needs to be noted that interpretation bias should be carefully considered and removed to the greatest extent possible.

c) Objective assessment

Objective task-related measures allow quantifying benefits of robot affect through such variables as accuracy, performance success, time it takes to complete the task, resource usage and others, depending on a particular task and scenario. One clear-cut advantage of this method is the removal, to a large extent, of both subject and interpretation bias. We can measure two types of performance this way:

- *Directly influenced by robot affect* – some affective phenomena are expected to provide task benefits regardless of whether any interaction is present. For example traits, in essence, suggest behavioral strategies optimized for certain types of tasks, environments and circumstances, and emotions help avoid pitfalls and attract attention to useful objects. This case is more straightforward and amenable to quantifiable assessment.
- *Indirectly influenced through participants' behavioral changes* – a person can change his/her actions in response to a robot's affective behaviors, and this, in its turn, can lead to a change in overall performance. One notable example of this is presented in a very cleverly designed study [11], in which the authors measured task performance that changed as a result of a robot's expression of anxiety during the scenario. In particular, as the robot's anxiety (expressed by voice) increased, the participants were alerted to the impending deadline, and worked more efficiently.

We believe that significant effort should be placed into developing objective measures for affective behavior assessment, especially given that the use of this method is almost non-existent in the current robot affect research (in a great part due to the challenges described earlier). Such measures would produce quantifiable and hopefully predominantly unbiased results.

IV. Conclusion

We are currently addressing the discussed open questions in humanoid emotions and affect by integrating the TAME framework within the Georgia Tech *MissionLab*[1] [6] software system and prototyping it on a Nao humanoid robot (Fig. 2). This robotic platform is small, but sufficiently expressive, and we hope it will prove an adequate test bed for implementing humanoid affect and exploring the challenges and opportunities it provides.

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

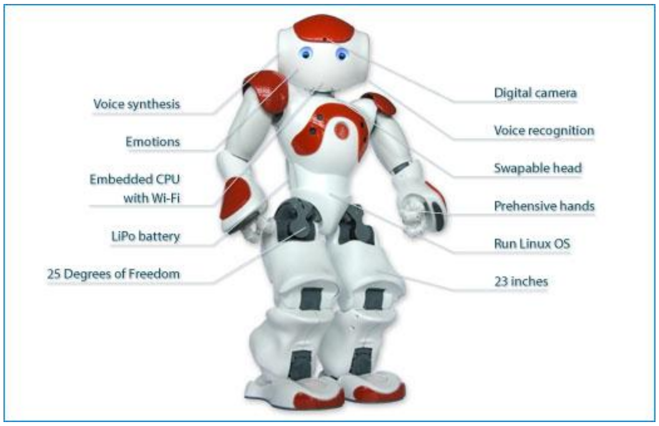

Figure 2: Nao Robot (source Aldebaran Robotics)

REFERENCES

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

[2] Moshkina, L., Arkin, R.C., *On TAMEing Robots.* Interanational Conference on Systems, Man and Cybernetics, 2003.

[3] Moshkina, L., Arkin, R.C. *Human Perspective on Affective Robotic Behavior: A Longitudinal Study*. IEEE IROS, 2005.

[4] Moshkina, L., *An Integrative Framework for Affective Agent Behavior*, IASTED International Conference on Intelligent Systems and Control, 2006.

[5] Arkin, R.C., *Behavior-based Robotics*, MIT Press, 1998.

[6] MacKenzie, D., Arkin, R.C., and Cameron, J., 1997 . "Multiagent Mission Specification and Execution", *Autonomous Robots*, Vol. 4, No. 1, Jan. 1997, pp. 29-57.

[7] Nass, C., Moon, Y., Morkes, J., Kim, E-Y., Fogg, B.J., *Computers are social actors: A review of current research*, in *Moral and ethical issues in human-computer interaction*, B. Friedman, Editor. 1997, CSLI Press: Stanford, CA.

[8] Saucier, G., *Mini-Markers: a Brief Version of Goldberg*ׄ*s Unipolar Big-Five Markers.* Journal of Personality Assessment, 1994. **63**(3): p. 506-516.

[9] Watson, D., Clark, L.A., Tellegen, A., *Development and validation of brief measures of positive and negative affect: The PANAS scales.* Journal of Personality & Social Psychology, 1998. **56**(6): p. 1063-1070.

[10] Morris, J.D., *Observations: SAM: The Self-Assessment Manikin*, *An Efficient Cross-Cultural Measurement of Emotional Response*, Journal of Advertising Research, November 1995.

[11] Lang, P.J., Bradley, M.M., and Cuthbert, B.N., *International Affective Picture System (IAPS): Technical Manual and Affective Ratings*, University of Florida, Center for Research in Psychophysiology; Gainesville: 1999

[12] Scheutz, M.; Schermerhorn, P.; Kramer, J.; and Middendorff, C., *The utility of affect expression in natural language interactions in joint human-robot tasks*, ACM HRI, 226–233.