Toward Affective Cognitive Robots for Human-Robot Interaction

M. Scheutz, J. Kramer, C. Middendorff, P. Schermerhorn, M. Heilman, D. Anderson, P. Bui
Artificial Intelligence and Robotics Laboratory
Department of Computer Science and Engineering
University of Notre Dame, Notre Dame, IN 46556, USA
{mscheutz,jkramer3,cmidden1,pscherm1,mheilman,danders,pbui}@cse.nd.edu

Abstract

We present a brief overview of an architecture for a complex affective robot for human-robot interaction.

Implementation of an Affective Architecture

Affect might have several functional roles in agent architectures (Scheutz 2004). Some have argued that affect serves the purpose of integration and management of multiple processes and is thus required for the effective functioning of an autonomous system (Ortony, Norman, & Revelle forthcoming). Specifically, affect allows for motivational signals originating not from changes in the external environment detected *via* sensors, but from components within the architecture itself (e.g., from deliberative subsystems). Such signals can then influence various other parts of the architecture and modify goal management, action selection, and learning.

Figure 1 depicts a partial view of the functional organization of the proposed affective architecture for complex robots.[1] Columns separate sensors, perceptual, central, and action processing components, and effectors. All boxes depict autonomous computing components that can operate in parallel and communicate via several types of communication links (in the figure, only links that are part of affective processing pathways are shown).[2] Labels of components denote the functional role of the component in the overall system. The colored/grayish components are part of the affect system, which is used to control the robot's actions and drive the behavior of the robot over time by providing internal feedback about the success or failure of an action.

The implementation of the proposed architecture is–almost by necessity–work in progress given the intrinsic complexity of the system.[3] However, several subsystems are functional and their interactions allow the robot to integrate sensory information from sonar, laser, bumper, vision, and auditory sensors, process information at several levels in parallel, and perform several actions: from finding and tracking multiple people in a room, to following trajectories in a room, to producing emotional speech output. For space reasons, we can only briefly touch on some subsystems.

The *vision subsystem*, for example, provides information about detected faces using and histogram methods from (Yang, Kriegman, & Ahuja 2002) as well as skin color, color of clothes, and camera angle relative to the robot, which are used in conjunction with information about the distance of an object (coming from sonar and laser sensors) to find and track people in a room (Scheutz, McRaven, & Cserey 2004) and determine some of their salient features such as height for future identification. The emotion recognition component currently only classifies faces as "happy", "sad" or "neutral".

The *natural language processing subsystem* integrates and extends various existing components (CMU's *SPHINX* and IBM's *ViaVoice* for spoken word recognition, an enhanced version of "thought treasure" (Mueller 1998) for natural language understanding and production, and a modified version of the University of Edinburgh's *Festival* system for speech synthesis).

The *action control subsystem* is based on a novel *action interpreter*, which interprets scripts for natural language understanding augmented by action primitives for the control of actions (see following Subsection). These scripts can be combined in hierarchical and recursive ways, yielding complex behaviors from basic behavioral primitives, which are grounded in basic *skills* (the bottom layer control structures are implemented as motor schemas based on reactive deliberative GLUE integration component (Kramer & Scheutz 2003). Moreover, several spatial maps are used for the representation of locations of the robot, people, and other salient objects in the environment, as well as path planning and high-level navigation.

All motions are controlled by the action interpreter based on scripts (like the following "serve-drink" script for a robotic waiter) are stored in long-term memory:

```
====serve-drink//serve.V
role01-of=waiter|
role02-of=human|
role03-of=beverage|
```


[1] This diagram of the architecture is restricted to visual and auditory sensors and camera motors and speaker as effectors; other sensors such as sonar and laser sensors, or effectors, such as grippers and wheel motors, are not shown.

[2] The implementation builds on the ADE system available at `http://ade.sourcefore.net/`.

[3] Most components are implemented in some rudimentary form, even though the majority of the components does not yet update in parallel.

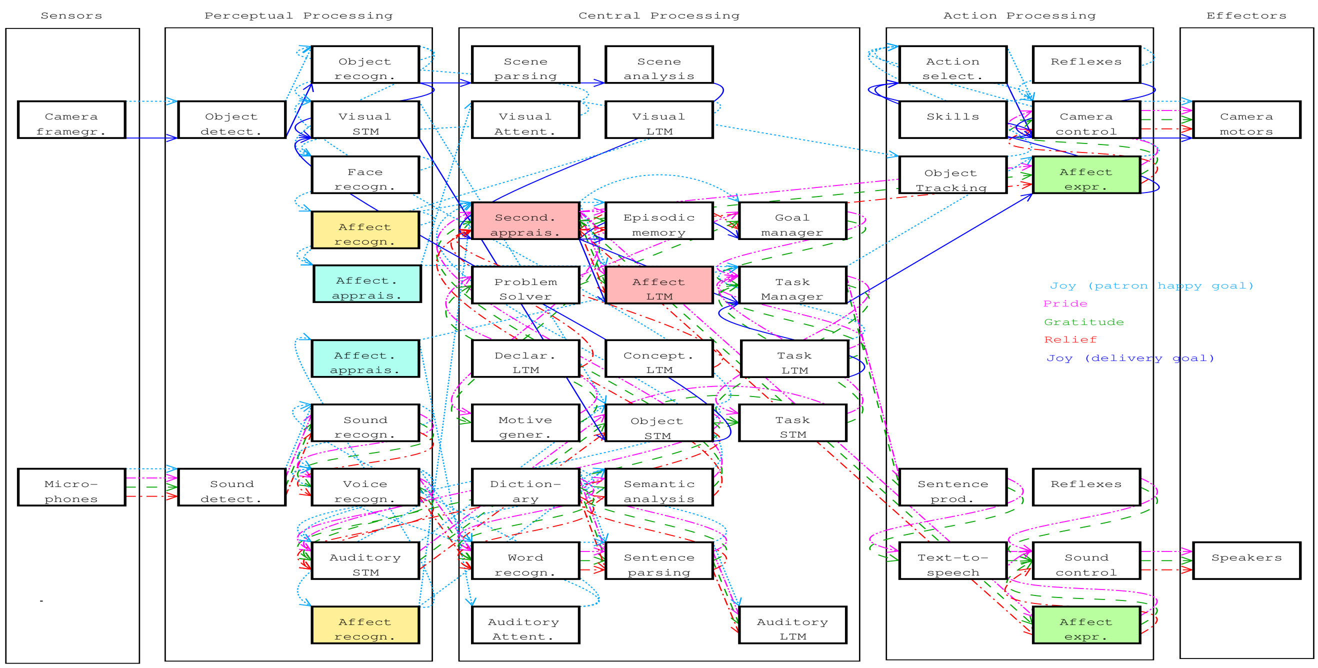

Figure 1: A partial view of the proposed affect architecture for a complex robot. Boxes depict concurrently running components of varying complexity and colored/grayish components indicate affective control mechanisms. The five different arrow types depicted indicated the information flow through the architecture involved in processing five different affective processes.

```
role04-of=bar|
timeout-of=600sec|
event01-of=[wander waiter]|
event02-of=[shiftFOA waiter human]|
event03-of=[converse waiter human]|
event04-of=[say-to waiter human [fetch-from waiter bar beverage]]|
event05-of=[move-to waiter bar]|
event06-of=[say-to waiter bar [active-goal waiter
                                [fetch-from waiter bar beverage]]]|
event07-of=[move-to waiter human]|
event08-of=[shiftFOA waiter human]|
event09-of=[say-to waiter human [active-goal-weaker waiter
                              [fetch-from human waiter beverage]]]|
```

"Roles" indicate entities and "events" denote of descriptions of actions and state-of-affairs involving these entities. In the above script all events denote actions (MOVE-TO, SHIFT-FOA for "shift focus of attention", and SAY-TO) are action primitives, while "wander" and "converse" refer to other scripts, leading to a recursive structure of scripts).

Note that the same script can also be used for "understanding" actions of others without having to carry them out, because the "serve-drink script" is not defined in an *indexical way* (i.e., assuming that the robot is the waiter), but rather defined generically so that the action interpreter can use the present context to determine who should assume the waiter role. If based on the context an entity other than the robot should be the waiter, then this will only lead to a new binding of the symbol "waiter" in the present context in the robot's working memory, whereas a binding of "waiter" to "self" will cause the robot to perform an action.

The architecture has been implemented on a fully autonomous Pioneer Peoplebot from ActivMedia (with sonar, pan-tilt-zoom camera, a SICK laser range finder, three sonar rings, two microphones, two speakers, one built-in Linux PC104 board, and two added Dell laptops, one running Linux, the other running Windows XP, connected through a local Ethernet network, with wireless access to the outside world).

References

Kramer, J., and Scheutz, M. 2003. GLUE - a component connecting schema-based reactive to higher-level deliberative layers for autonomous agents. In Weber, R., ed., *Proceedings of the 16th International FLAIRS Conference*, 22–26. AAAI Press.

Mueller, E. T. 1998. *Natural language processing with ThoughtTreasure*. New York: Signiform.

Ortony, A.; Norman, D.; and Revelle, W. forthcoming. Effective functioning: A three level model of affect, motivation, cognition, and behavior. In Fellous, J., and Arbib, M., eds., *Who needs emotions? The Brain Meets the Machine*. New York: Oxford University Press.

Scheutz, M.; McRaven, J.; and Cserey, G. 2004. Fast, reliable, adaptive, bimodal people tracking for indoor environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Scheutz, M. 2004. Useful roles of emotions in artificial agents: A case study from artifici al life. In *Proceedings of AAAI 2004*.

Yang, M.-H.; Kriegman, D. J.; and Ahuja, N. 2002. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(1):34–58.