# Development of a Robot with a *Sense of Self*

**K. Kawamura, W. Dodd, P. Ratanaswasd and R. A. Gutierrez**
*Center for Intelligent Systems, Vanderbilt University*[1]
*Nashville, Tennessee 37235-0131 USA*
*kawamura@vuse.vanderbilt.edu*

***Abstract* – This paper describes our efforts to develop a robot with *a sense of self* using a multiagent-based cognitive architecture and control with three distinctive memory systems, namely (1) spatio-temporal short-term memory, (2) procedural / declarative / episodic long-term memory and (3) a task-oriented adaptive working memory. Such a robot may be called a *cognitive robot*. Cognitive robots share a number of key features with conscious machines. We are exploring the interface between cognitive robots and machine consciousness through an internal model called the Self Agent.**

***Index terms* –** cognitive robot, cognitive control, machine consciousness, self agent, adaptive working memory

## I. INTRODUCTION

In recent years, design philosophies in the field of robotics have followed the classic dialectic. Initial efforts to build robots capable of perceiving and interacting with the world around them involved explicit knowledge representation schemes and formal techniques for manipulating internal representations. Tractability issues gave rise to antithetical approaches, in which deliberation was eschewed in favor of dynamic interactions between primitive reactive processes and the environment [1]. Many studies have shown the need for both, motivating work towards hybrid architectures [2]. While such an integration of robotic body, sensor and artificial intelligence (AI)-based software offers the promise of robots which are fluent in sensorimotor operations and capable of adjusting their behavior in different situations, the reality is quite different from what researchers had hoped.

Most robots currently can perform only those or similar tasks which they were programmed for and very little emerging behaviors are exhibited. What is needed is an alternative paradigm for behavior learning and task execution. We believe that robust and timely responses to the full range of contingencies often present in complex task environments will require something more than the combination of traditional approaches. Specifically, we see our brain's cognitive flexibility and adaptability as desirable design goals for a next generation of intelligent robots. Several cognitive architectures have been implemented for the purpose of testing human psychological models [3][4], but such models have not been applied to robotics.

This new generation of robots should be able to recognize and deal with situations in which its traditional reactive and reasoning abilities fall short of meeting complex task demands.

At ICAR2003 in Coimbra, Portugal, we proposed a concept of a cognitive robot [5] as a robot that knows what it is doing and reflects on past experience to deal with new situations. In the current paper, we describe further details of such a cognitive robot architecture for our humanoid robot ISAC [6] with three distinctive memory structures: short-term and long-term memories and a working memory system. Short-term memory is a sparse data structure called the Sensory EgoSphere (SES) [7] containing spatio-temporal sensory data acquired within a recent time frame. Long-term memory is composed of behaviors, semantic knowledge, and past experience. A working memory system allows the robot to focus attention on the most relevant features of the current task and provide robust operation in the presence of distracting irrelevant events [8][9].

---

[1] This work is supported in part under NSF grant EIA0325641, "ITR: A Biologically Inspired Adaptive Working Memory System for Efficient Robot Control and Learning".

## II. MULTIAGENT-BASED COGNITIVE ROBOT ARCHITECTURE

A humanoid is an example of a robot that requires intelligent behavior to act with generality in its environment. As the complexity of a task grows, so do software complexities necessary to process sensory information and to control actions purposefully. The development and maintenance of complex or large-scale software systems can benefit from domain-specific guidelines that promote code reuse and integration through software agents. Information processing in ISAC, from perception through action execution, is integrated into a multiagent-based software architecture based on the Intelligent Machine Architecture (IMA) [10]. The IMA was designed to provide such guidelines and allows for the development of subsystems capable of environmental modeling and robot control through the collections of IMA agents and associated memories, as shown in Figure 1 [11].

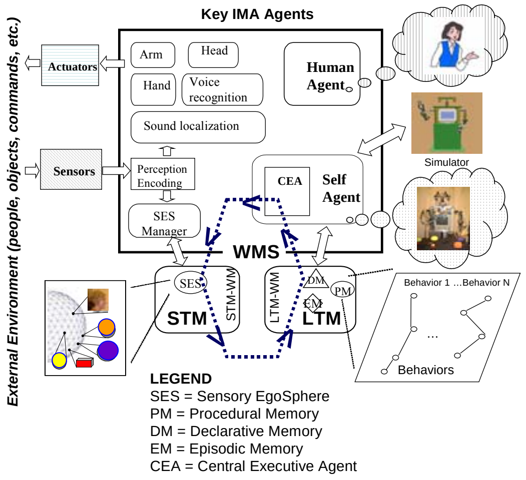

Fig. 1. Multiagent-based cognitive robot architecture.

## III. MACHINE CONSCIOUNESS AND THE SELF AGENT

### A. Machine Consciousness

A considerable amount of research is being conducted in the area of consciousness in humans and machines, e.g., see [12][13][14]. Some argue that machine consciousness is not only possible but also inevitable for that to happen [15][16][17].

Before explaining what human consciousness is, however, "it may be advantageous to build a machine which can demonstrate key functions of a conscious agent so that many questions will be clarified" [15]. For example, we are working on a software agent called the *Self Agent* that uses emotion, attention and cognitive control to model "consciousness" for ISAC.

### B. *Self Agent*

Our initial attempt to develop machine consciousness in ISAC is through a set of tightly-coupled atomic agents trying to achieve a common goal. This concept was inspired by Minsky's work in the *Society of Mind* [18].

The Self Agent (SA) represents the *sense of self* [19] through monitoring the robot's own internal state as well as the progress of task execution via sensor signals, agent communications and working memory. The internal representation of the robot's self should continually be updated and enhanced to allow the system to reason and act based on its status and the context of assigned tasks.

The SA also responds to commands given by humans through the Human Agent and is responsible for controlling task execution. Figure 2 illustrates the current design of the Self Agent and its interaction with other components. So far, the Intention Agent, the Pronoun Agent, and the Description Agent have been implemented [11]. We are currently developing the Central Executive Agent (CEA) and the Emotion Agent (See Sect. V).

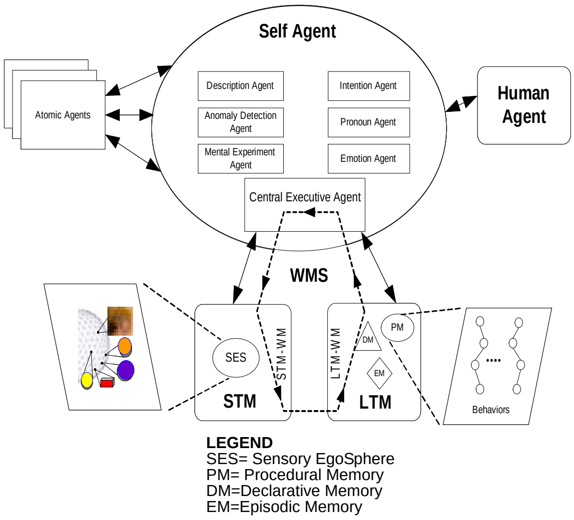

Fig. 2: Current Design of the Self Agent.

## IV. MEMORY STRUCTURE

ISAC's memory structure is divided into three classes: Short-Term Memory (STM), Long-Term Memory (LTM), and the Working Memory System (WMS). The STM holds sensory information about the current environment, while the LTM holds learned and taught behaviors, semantic knowledge, and past experience. The WMS holds task-specific STM and LTM information and streamlines the information flow to the cognitive processes during the task as described in section C.

### *A. Short-Term Memory: The Sensory EgoSphere*

We are using a sparse sensory data structure called the Sensory EgoSphere (SES) to hold STM data. The SES, inspired by the egosphere concept defined by Albus [20], serves as a spatio-temporal STM for a robot [7]. The SES is structured as a geodesic sphere centered between the robot's cameras and indexed by azimuth and elevation. Each vertex of the geodesic sphere contains a database representing a detected stimulus at the corresponding angle (Figure 3).

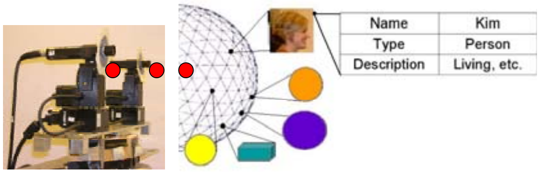

Fig.3. Structure of the Sensory EgoSphere.

Memories in the SES can be retrieved by angle, stimulus content such as key words or colors, or time of posting. This flexibility in searching allows for easy memory management, posting, and retrieval.

### *B. Long-Term Memory: Procedural, Episodic, and Declarative Memories*

Long-term memory (LTM) is divided into three types: Procedural, Episodic, and Declarative. LTM stores information such as *skills learned* and *experiences gained* for future retrieval. Procedural Memory (PM) holds motion primitives and behaviors needed for movement, such as how to *reach to a point*. Behaviors are derived using the spatio-temporal Isomap method proposed by Jenkins and Mataric [22]. A short description of how it operates follows: Motion data are collected from teleoperation then segmented into a set of motion primitives. Then spatio-temporal Isomap dimension reduction, clustering and interpolation methods are applied to the motion segments to produce motion primitives and behaviors [23] (Figure 4).

Motion skills for each behavior must be interpolated in order to be used in specific situations. The interpolation method we are using is the Verbs and Adverbs method developed in [24]. This technique describes a motion (verb) in terms of its parameters (adverbs) that generates a new movement based on the similarity of stored motions.

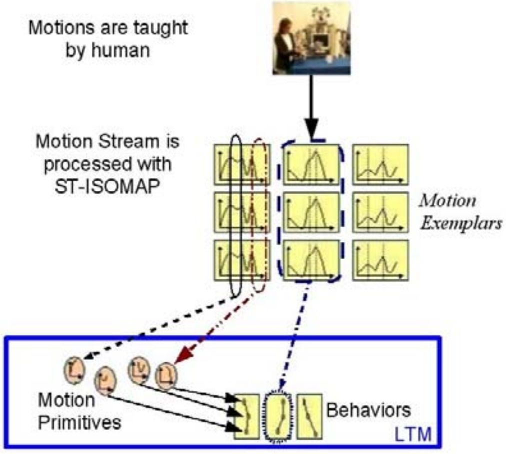

Fig. 4. Derivation of Procedural Memory through human-guided motion stream.

Declarative Memory (DM) is a data structure about objects in the environment. Goals and task sequences are stored as DM units. After this is accomplished, an episodic memory will be developed to store past experience.

### *C. Working Memory System*

There is much evidence for the existence of working memory in primates [25][26]. Such a memory system is said to be closely tied to task learning and execution [27][28].

Working memory "represents a limited-capacity store for retaining information over the short term and for performing mental operations on the contents of this store" [28]. Inspired by this, we are investigating the utility of integrating the (adaptive) working memory structure into a robot in order to provide the embodiment necessary for exploring the critical issue of task learning.

Our hypothesis is that this integration will lead to a more complex, but realistic robotic learning system involving perceptual systems, actuators, reasoning, attention, emotion, and short- and long-term memory structures [9].

It is interesting to note that Just and Carpenter of Carnegie Mellon University referred to working memory as "the blackboard of the mind" [29] since our earlier ISAC software development was based on a "blackboard" architecture [30].

The Working Memory System (WMS) we are developing includes the Central Executive Agent (See Section V) and short- and long-term working memories.

## V. COGNITIVE CONTROL AND THE CENTRAL EXECUTIVE AGENT

### *A. Cognitive Control*

Cognitive control in humans is the ability to "consciously manipulate thoughts and behaviors using attention to deal with conflicting goals and demands" [31]. As levels of human behavioral processes range from reactive to full deliberation, cognitive control must be able to switch between these levels to cope with the demand of task and performance, particularly in novel situations. According to a number of cognitive psychologists, cognitive control in human is performed through the working memory in the pre-frontal cortex (PFC) [8][28]. Furthermore, attention and emotion play an important role in human's decision and task execution [32].

Inspired by these concepts, we have implemented cognitive control in ISAC using a mechanism called the Central Executive Agent.

### *B. Central Executive Agent*

ISAC's cognitive control is modeled and implemented based on Baddeley and Hitch's psychological human working memory model [27]. Their model consists of the "central executive" which controls two working memory systems, i.e., phonological loop and visuo-spatial sketch pad. Cognitive control in ISAC is implemented using the Central Executive Agent (CEA) that interfaces with the WMS.

The CEA's functions include task planning, action selection and action execution. Figure 5 illustrates the interaction between the CEA and the WMS during an action selection and execution process implemented for ISAC [33].

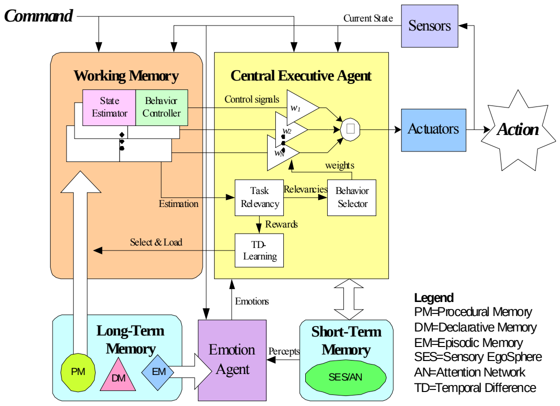

Fig. 5. Interaction between the CEA and WMS during a task execution.

Goal-oriented behavior selection and execution is being done in a modular fashion. Upon receiving a command, the CEA associates a set of behaviors based on past experience and places them in the WMS. Currently, past experience is stored and maintained in the form of teleoperated behaviors.

State estimators produce estimated states to calculate task relevancies of each behavior according to the goal. The behavior selector computes time-varying weights $w_i$ based on task relevancies to combine behaviors to generate the final action. Results from the task execution are used to calculate expected rewards for the TD-Learning [34].

Figure 6 shows an example of time-varying weight distribution among three behaviors during a simulation where three behaviors were selected and combined to achieve the goal *"point to"* a position on a table (Figure 7) [33].

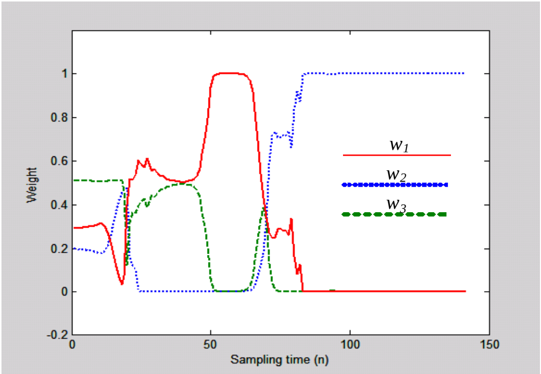

Fig. 6. Weight distribution among selected behaviors during a *"point to"* action.

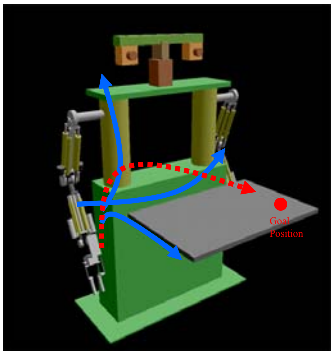

Fig. 7. Motions of stored behaviors (solid lines) and combined behavior motion (dotted line).

### *C. Emotion Agent*

Implementing emotion-based sensor signal processing in robotics is becoming popular [35]. Meantime, the interaction of attention and emotion in the human brain is increasingly well understood [32]. Emotional state of the robot describes what the robot feels toward the task and the environment based on past experience.

Inspired by this, we are adding an Emotion Agent to the Self Agent to conduct cognitive control experiments. Section VI describes the current emotion-based cognitive control experiment using the emotion *fear*.

## VI. CURRENT COGNITIVE CONTROL EXPERIMENT

We have designed an integrated cognitive system experiment based on the CEA, attention, emotion and the adaptive working memory system as follows:

1. ISAC is trained to learn specific object using voice, vision, attention *(Learn by association)*
2. ISAC is asked to point to one of the learned objects *(Use of short-term memory of the object and long-term procedural memory)* (Figure 8)
3. ISAC is asked to visually track the object held by a human *(Color tracking)*
4. A person enters the room and yells "Fire!" ISAC using attention, emotion and cognitive control, suspend the current tracking task and warn everyone to exit the room *(Cognitive control).*

Steps 1-3 have already been implemented and presented elsewhere [21][36]. In Step 4, ISAC's cognitive control must

- Pay attention to new stimulus
- Use emotion to activate cognitive control.

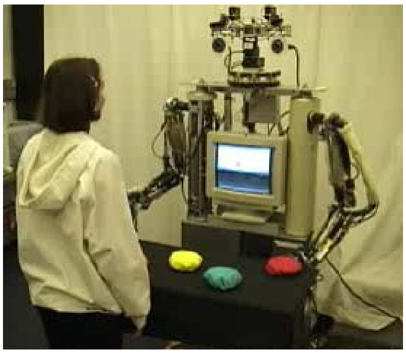

Fig. 8. ISAC is asked to point to one of the learned objects.

This cognitive control experiment is being done through integrating the working memory and cognitive control with the existing IMA agents as shown in Figure 10.

This experiment tries to demonstrate that "The artificial cognitive machine is not governed by any programs and therefore will not execute any preprogrammed decision commands like the IF-THEN ones" [p.216, 15].

Fig. 9. Cognitive Control Experiment.

## VII. CONCLUSIONS

Realization of general-purpose robots with adult-level intelligence continues to be the dream of many robotic researchers. During the past decade, we have seen major advances in the integration of intelligent robots and expect this trend to continue.

The next grand challenge will be in the integration of body and mind. This paper described our efforts towards this challenge through the realization of a cognitive robot using cognitive control, attention, emotion, and an adaptive working memory system. Our multiagent-based cognitive approach is an attempt to capture brain-style computation without necessarily committing to the neural-level details.

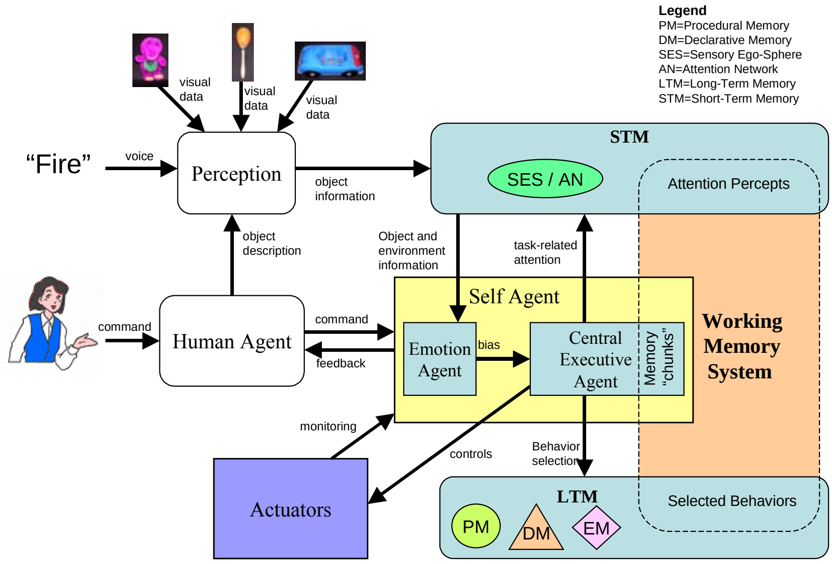

Fig. 10. Interactions among IMA agents and memories during the cognitive control experiment.

## REFERENCES

[1] R.A. Brooks, "Intelligence without representation," *Artificial Intelligence* vol.47 nos.1-3, pp.139-160, 1991.

[2] E. Gat, "Three level architectures," Chapter 8 of *Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems* (E. Kortenkamp, R.P. Barasso and R. Murphy, Eds.), AAAI Press, pp. 195-210, 1998.

[3] J.R. Anderson, D. Bothell, M.D., Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of the mind", (ACT-R), *Psychological Review* vol. 111, no. 4. pp.1036-1060, 2004.

[4] J.E. Laird, A. Newell, and P.S. Rosenbloom, "SOAR: An architecture for general intelligence", *Artificial Intelligence,* vol. 33, no.1, pp.164, 1987.

[5] K. Kawamura, D.C. Noelle, K.A. Hambuchen, and T.E. Rogers, "A multi-agent approach to self-reflection for cognitive robots", *Proc. of 11th Int'l Conf. on Advanced Robotics*, Coimbra, Portugal, June 30 - July 3, 2003, pp. 568-575, 2003.

[6] K. Kawamura, R.A. Peters II, D.M. Wilkes, W.A. Alford, and T.E. Rogers, "ISAC: foundations in human-humanoid interaction," *IEEE Intelligent Systems,* July/August 2000, pp. 38-45, 2000.

[7] R.A. Peters II, K.A. Hambuchen, K. Kawamura, and D.M. Wilkes, "The sensory egosphere as a short-term memory for humanoids," *Proc. of the IEEE-RAS Int'l Conf. on Humanoid Robots*, Waseda University, Tokyo, Nov. 22-24, pp.451-459, 2001.

[8] R. O'Reilly, T. S. Braver, and J. D. Cohen, "A biologically based computational model of working memory", *Models of Working Memory: Mechanisms of active maintenance and executive control*, (A. Miyake and P. Shah, Eds.) Cambridge: Cambridge Univ. Pr., 1999.

[9] M. Skubic, D. Noelle, M. Wilkes, K. Kawamura, and J.M. Keller, "A biologically inspired adaptive working memory for robots," *AAAI Fall Symp., Workshop on the Intersection of Cognitive Science and Robotics: From Interfaces to Intelligence*, Washington DC, October 2004.

[10] R.T. Pack, D.M. Wilkes, and K. Kawamura, "A software architecture for integrated service robot development," *Proc. of IEEE Systems, Man and Cybernetics*, October 1997, pp. 3774-3779, 1997.

[11] K. Kawamura, R.A. Peters II, R. Bodenheimer, N. Sarkar, J. Park, A. Spratley, and K. A. Hambuchen, "Multiagent-based cognitive robot architecture and its realization," *Int'l Jo. of Humanoid Robotics*, vol. 1, no. 1, pp. 65-93, 2004.

[12] J.G. Taylor, "Paying Attention to Conscious- ness", *Trends in Cog. Sciences*, vol. 6, pp. 206-210, 2002.

[13] B. J. Baars, *In the Theater of Consciousness: The Workspace of the Mind*, Portland, OR: Book News, Inc., 2004.

[14] S. Franklin, “IDA: A Conscious Artifact?” Institute for Intelligent Systems, The Univer- sity of Memphis. *J. of Consciousness Studies*, vol.10, pp.47-66, 2003.

[15] P. O. Haikonen, *The Cognitive Approach to Conscious Machines*, Charlottesville, VA: Imprint Academic, March 2003.

[16] O. Holland, ed., *Machine Consciousness*, Charlottesville, VA: Imprint Academic, 2003.

[17] I. Aleksander, *How to Build a Mind. Toward Machines with Imagination,* Columbia University Press, NY, 2001.

[18] M. Minsky, *The Society of Mind*, NY: Simon & Schuster, 1985.

[19] R. Sanz, “Modeling, self and consciousness: further perspectives of AI research,” *Performance Metrics for Intelligent Systems Workshop (PerMIS)*, Aug. 13-15, 2002, NIST, Washington, DC, 2002.

[20] J.S. Albus, “Outline for a theory of intelligence,” *IEEE Trans Systems, Man, and Cybernetics*, vol. 21, no.3, pp.473–509, 1991.

[21] K.A. Hambuchen, Multi-Modal Attention and Binding using a Sensory EgoSphere, Ph.D. Dissertation, Nashville, TN: Vanderbilt University, May 2004.

[22] O.C. Jenkins and M.J. Matarić, “Automated derivation of behavior vocabularies for autonomous humanoid motion,” *2nd International Joint Conference on Autonomous Agents and Multiagent Systems*, 2003.

[23] D. Erol, J. Park, E. Turkay, K. Kawamura, O.C. Jenkins and M.J. Mataric, “Motion generation for humanoid robots with automatically derived behaviors,” *Proc. of IEEE Int’l. Conf. on Systems, Man, and Cybernetics,* Washington, DC, Oct. 6-8, 2003, pp. 1816-1821, 2003.

[24] C. Rose, M.F. Cohen, and B. Bodenheimer, “Verbs and adverbs: Multidimensional motion interpolation”, *IEEE Computer Graphics and Appl.,* vol. 18, no. 5, Sept-Oct 1998, pp. 32-40, 1998.

[25] S. Funahashi and K. Kubota, “Working memory and prefrontal cortex”, *Neuroscience Research,* vol. 21, pp. 1-11, 1994.

[26] E.K. Miller, C.A. Erickson, and R. Desimone, “Neural mechanisms of visual working memory in prefrontal cortex of the macaque”, *Jo. of Neuroscience,* vol. 16, pp. 5154-6, 1996.

[27] A.D. Baddeley, *Working Memory*. Oxford: Clarendon Press, 1986.

[28] M.S. Gazzaniga, R.B. Ivry, and G.R. Mangun, *Cognitive Neuroscience: The biology of the mind, 2nd ed.*, New York: W.W. Norton & Company, p. 311, 2002.

[29] P.S. Goldman-Rakic, Working Memory and the Mind, *Scientific American,* pp. 11-117, September 1992.

[30] K. Kawamura, R.A. Peters II, S. Bagchi, M. Iskarous, and M. Bishay, “Intelligent Robotic Systems in Service of the Disabled”, *IEEE Trans. on Rehabilitation Engineering,* Vol. 3, No. 1, pp. 14-21, March 1995.

[31] T. S. Braver and J. D. Cohen, “On the control of control: The role of dopamine in regulating prefrontal function and working memory,” In S. Monsell & J. Driver, eds., *Control of Cognitive Processes: Attention and Performance XVIII,* Cambridge, MA: MIT Press, pp. 713-738, 2000.

[32] J. G. Taylor, “Paying attention to consciousness,” *Progress in Neurobiology* (Elsevier), vol. 71, pp.305-335, 2003.

[33] P. Ratanaswasd, W. Dodd, K. Kawamura, and D. Noelle, “Modular behavior control for a cognitive robot,” *12th Int’l Conf. on Advanced Robotics (ICAR)*, Seattle WA, July 18-20, 2005, *in review.*

[34] R. S. Sutton, “Learning to predict by the method of temporal differences,” *Machine Learning,* vol.3, pp.9-44, 1988.

[35] C. Breazeal, *Designing Social Robots*, MIT Press, 2002.

[36] J. Rojas, *Sensory Integration with Articulated Motion on a Humanoid Robot,* Master’s Thesis, Nashville, TN: Vanderbilt University, May 2004.