Development of a Robot with a *Sense of Self*

K. Kawamura, W. Dodd, P. Ratanaswasd and R. A. Gutierrez
Center for Intelligent Systems[1]
Vanderbilt University
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 based on psychological and computational neuroscience models. Such a robot may be called a *cognitive robot.* Cognitive robots share a number of key features with conscious machines. We explore the interface between cognitive robots and conscious machines through an internal model called the Self Agent.**

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

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] [2]. Many studies have shown the need for both, motivating work towards hybrid architectures [3]. 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 hoped for. Most robots currently can perform only those or similar tasks for which they were programmed for and very little emerging behaviors are exhibited. What we need is an alternative paradigm to traditional AI (both *hard* and *soft*)-based approach for behavior learning. 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. 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 [4] as a system which knows what it is doing and reflect on past experiences to deal with new situations. In the current paper, we propose more details of such a cognitive robot architecture for our humanoid robot ISAC [5] with three distinctive memory structures: short-term and long-term memories and a working memory system. Short-term memory is a data structure called the Sensory EgoSphere (SES) [6] and contains spatio-temporal sensory data acquired within a recent time frame. Long-term memory (LTM) is composed of behaviors and a set of

[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".

semantic knowledge for learned or taught tasks. 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 [7][8].

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. Especially in interactions with humans, the robot must be able to adapt its behaviors to accomplish goals timely and safely. As the complexity of interaction grows, so grows the complexity of the software necessary to process sensory information and to control action purposefully. The development and maintenance of complex or large-scale software systems can benefit from domain-specific guidelines that promote code reuse and integration. Information processing in our humanoid robot ISAC, from perception through action execution, is integrated into a multiagent-based software architecture based on the Intelligent Machine Architecture (IMA) [9]. 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. Within this cognitive architecture, the Human Agent and the SES represent the humans and objects in the external environment, while the Self Agent, the LTM, and the working memory system (WMS) provide internal sense of self and behavioral and cognitive control mechanisms [10][11].

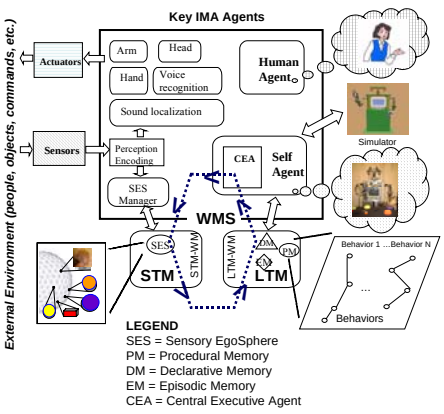

Fig. 1. MultiAgent-based cognitive architecture.

III. MACHINE CONSCIOUNESS AND THE SELF AGENT

A. Machine Consciousness

Researchers desire to build an artificial brain. As we enter the new century, a considerable amount of research is being conducted in the area of consciousness in humans and machines [12][13]. An increasing number of scientists today agree that is not only possible but also inevitable for that to happen [14][15][16][17].

Before fully explaining what human consciousness is, it may be advantageous to build a machine which can demonstrate key functions of a conscious agent so that many questions will be clarified [12]. For example, we are working on the Self Agent which uses emotions, attentions and cognitive control to deal with new situations. A similar multi-agent approach to describe consciousness has been proposed by Franklin [14].

A mere input-output reproduction of human-like conscious actions does not imply true consciousness in robot, but we expect to get a better understanding of consciousness as we gain more insight with the machine consciousness. We think other human beings to be conscious only because we ourselves are conscious, but, to truly know, we would have to get into their minds. The same could be true for conscious machines.

B. Self Agent

Our initial attempt to develop machine consciousness in ISAC is through the Self Agent consisting of 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* 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 [19]. The SA also responds to commands given by humans through the HA 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 [10]. We

are currently developing the Central Executive Agent (CEA) and the Emotion Agent. Details of the CEA and its relation to the Emotion Agent will be described in Section V.

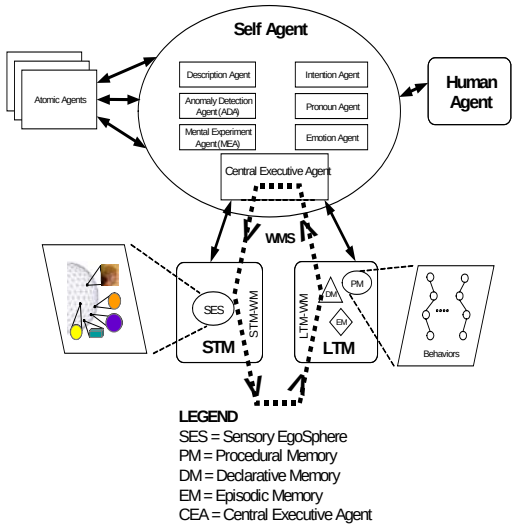

Fig. 2: Current Design of the Self Agent.

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 information about the current environment while the LTM holds learned behaviors, semantic knowledge, and past experience, i.e., episodes. The WMS holds task-specific STM and LTM information and streamlines the information flow to the cognitive processes during the task as detailed in section IV-C.

A. Short-Term Memory: The Sensory EgoSphere

Currently, we are using a structure called the Sensory EgoSphere (SES) to hold STM data. The SES is a data structure inspired by the egosphere concept as defined by Albus [20] and serves as a spatio-temporal short-term memory for a robot [7]. The SES is structured as a geodesic sphere that is centered at a robot's origin and is indexed by azimuth and elevation.

The objective of the SES is to temporarily store exteroceptive sensory information produced by the sensory processing modules operating on the robot. Each vertex of the geodesic sphere can contain a database node detailing a detected stimulus at the corresponding angle (Figure 3)..

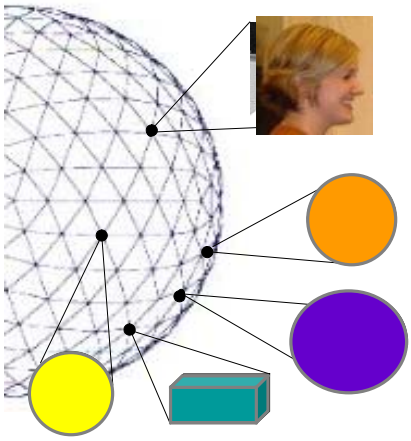

Fig.3. Structure of the Sensory EgoSphere [21].

Memories in the SES can be retrieved by angle, stimulus content, or time of posting. This flexibility in searching allows for easy memory management, posting, and retrieval. (See Section IV.C.)

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

LTM is divided into three types: Procedural Memory, Episodic Memory, and Declarative Memory. Like that in a human brain, the LTM stores information such as *skills learned* and *experiences gained* in the long term for future retrieval.

The part of the LTM called the Procedural Memory (PM) [22] 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 Mataríc [23]. A short description of how it operates is shown in figure 4.

Motion data are collected from the teleoperation of ISAC. The motion streams collected are then segmented into a set of motion primitives. The central idea in the derivation of behaviors from motion segments is to discover the spatio-temporal structure of a motion stream. This structure can be estimated by extending a nonlinear dimension reduction method called Isomap [24] to handle motion data. Spatio-temporal Isomap dimension reduction, clustering and interpolation methods are applied to the motion segments to produce Motion Primitives (Figure 4). Behaviors are formed by further application of the spatio-temporal Isomap method and linking Motion Primitives with transition probabilities [22].

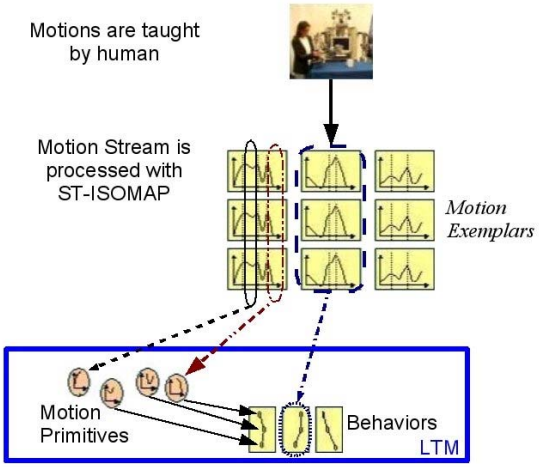

Fig. 4. Derivation of Procedural Memory through human-guided motion stream.

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 [25]. This technique describes a motion (verb) in terms of its parameters (adverbs) that allow ISAC to generate a new movement based on the similarity of stored motions.

Figure 5 depicts a current representation of the PM data structure. At the top of this structure, behavior descriptions will be stored which will allow us to identify what each behavior can contribute to solving a given motor task. Each entry in the behavior table will contain pointers to the underlying motion primitives.

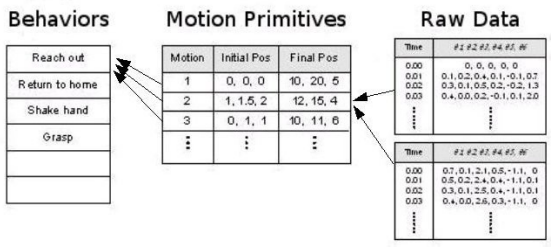

Fig. 5. Structure of Procedural Memory data unit.

Currently under development, the Episodic Memory (EM) system will hold records of past experiences in a time-indexed format. The purpose of the EM is to allow the cognitive process to trigger new action, selection, and execution. The Working Memory System (Sect. IV-C) records its own contents with reward information for the duration of a single task and posts this information to a LTM data unit.

The Declarative Memory (DM) currently is a data structure about objects in the environment. In the future, we plan to expand to include semantic knowledge.

C. *Attention and the Working Memory System*

There is much evidence for the existence of working memory in primates [26][27]. Such a memory system is closely tied to the learning and execution of tasks, as it contributes to decision-making capabilities by focusing on task-essential capabilities and information, and by discarding distractions [28] [29] [30].

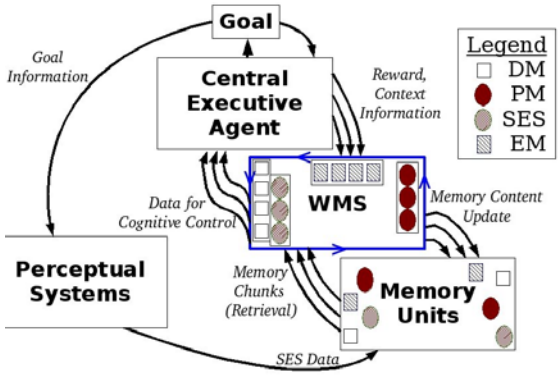

Fig. 6. Structure of the Working Memory System.

Inspired by this, we are currently investigating the utility of a combination of psychological and computational neuroscience-based working memory system (WMS) to be used in our humanoid robot Using an attentional network, highly task-dependent information is stored in the WMS, along with primitive motion information stored in the LTM. Representations in the WMS are continually updated and can be internally refreshed even in the absence of stimuli. This relevant short-term information exists throughout the task duration.

The Sensory EgoSphere can facilitate the direction of attention to external sensory events [31]. Because sensory processors report all exteroceptive events to the SES, the attention network is able to search the SES for both task relevant sensory data and unexpected yet salient sensory data and registers them to the STM-SES.

As multiple events are registered in a common area, activation increases around a central node. Nodes that receive registration from task- or context-related events have their activations increased by the attention network. The attention network selects the node with the highest activation as the focus of attention. Sensory events that contributed to this activation are selected and those that fall within a specified time range of each other are passed into the WMS.

The design of the LTM-WMS is expected to take many different forms. The DM-WMS holds all information needed to complete the task and is filled by a simple keyword search. The EM-WMS is populated by an agent that selects EM units based on similarity to the current situation and emotional salience. EM units are generated by the contents of the whole WMS during task execution.

The PM-WMS holds behaviors needed to complete the current task, and serves to reduce the complexity of real-time error-driven execution of behaviors. We are testing these structures with the experiment outlined in Sect. VI.

V. COGNITIVE CONTROL AND THE CENTRAL EXECUTIVE AGENT

A. Cognitive Control

Cognitive control in human is the ability to consciously manipulate thoughts and behaviors using attention to deal with conflicting goals and demands [31] [32]. 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.

Cognitive robots should have the ability to handle unexpected situations and the ability to reason and learn to perform new tasks. According to cognitive psychologists, cognitive control in human is performed through the working memory in the pre-frontal cortex (PFC) [7][33][34].

Furthermore, attention and emotion play an important role in human's decision and task execution [35]. Inspired by this concept, 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 [30]. 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 Working Memory System. The CEA's functions include task planning, action selection and action execution.

Figure 7 illustrates the interaction between the CEA and the WMS during an action selection and execution process.

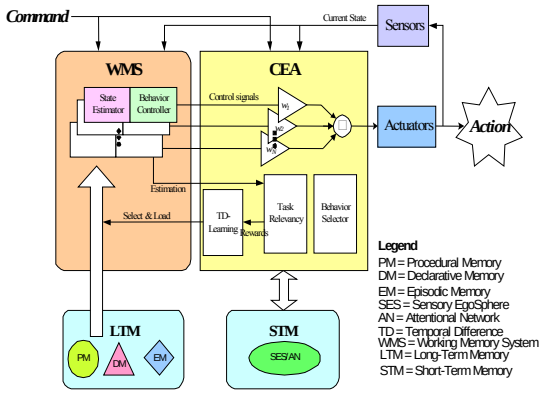

Fig. 7. Interaction between the CEA and WMS during task execution.

Action selection and execution within the CEA is done in a modular fashion as described in [36]. Upon receiving a command, the TD-Learning system [35] selects a set of behaviors based on past experience, and places them in the WMS. State estimators produce estimated states to calculate task relevancies for each behavior according to the assigned task. The CEA computes time-varying weights based on task relevancies and W_i to combine control signals to generate the final action. Figure 8 shows an example of weight distribution among three behaviors during one simulated action execution experiment [36].

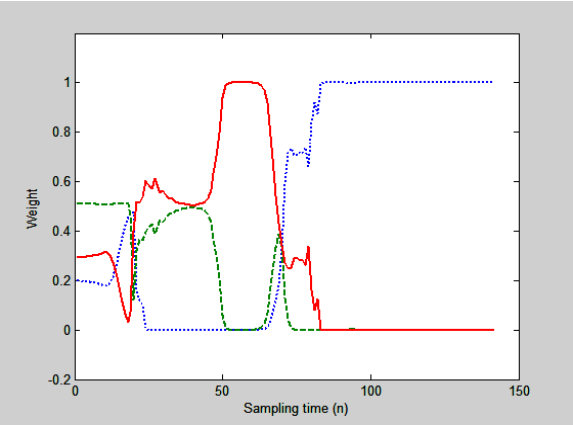

Fig. 8. Weight distribution among 3 behaviors within the WMS during an action execution.

Results from the task execution are used as rewards for the TD-Learning in selecting behaviors for a similar task in the future.

Implementing emotion-based sensor signal processing in robotics is becoming popular [37]. Meantime, the interaction of attention and emotion in the human brain is increasingly well understood [38]. 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 past episodic memory related to *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 and attention (Learning 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)
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][39]. In Step 4, ISAC's cognitive control must

- Pay attention to new stimulus
- Use emotion to activate the episodic memory
- Use EM to activate cognitive control

This cognitive control experiment is being done through integrating the WMS and cognitive control with the existing IMA agents as shown in figure 9.

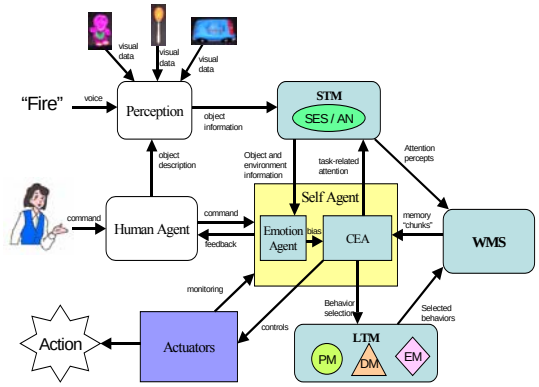

Fig. 9. Cognitive control experiment.

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, 12].

VII. CONCLUSIONS

Realization of general-purpose humanlike 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. It is our belief that such an approach offers the opportunity to help understand how to build human-like machines in the future.

ACKNOWLEDGMENT

The authors would like to thank Prof. David Noelle for his contribution for the discussion of working memory.

REFERENCES

[1] R. Arkin, *Behavior-Based Robotics*. Boston: MIT Press, 1998.

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

[3] 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.

[4] 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.

[5] 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.

[6] 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, 2001, pp 451-459, 2001.

[7] 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 University Press, 1999.

[8] 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.

[9] 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.

[10] 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.

[11] K. Kawamura, T.E. Rogers, K.A. Hambuchen and D. Erol, "Towards a human-robot symbiotic system," *Robotics and Computer Integrated Manufacturing*, vol. 19, pp. 555 – 565, 2003.

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

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

[14] S. Franklin, "IDA: A Conscious Artifact?" Institute for Intelligent Systems, The University of Memphis. *Journal of Consciousness Studies*, vol. 10, pp. 47-66, 2003.

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

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

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

[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] 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.

[23] 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.

[24] J.B. Tenenbaum, V. de Silva, and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction", *Science*, 290 (5500), pp. 2319–2323, 2000.

[25] 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.

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

[27] 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.

[28] R. O'Reilly, T. Braver, and J. 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 University Press, 1999.

[29] A.D. Baddeley, *Human Memory: Theory and Practice,* Oxford: Oxford University Press, 1990.

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

[31] M.M. Botvinick, T.S. Braver, D.M. Barch, C.S> Carter and D. Coh, "Conflict monitoring and cognitive control", *Psych. Rev.,* vol. 108, no.3, pp. 624-652, July 2001.

[32] C.M. MacLeod and P.W. Sheehan, "Hypnotic control of attention in the Stroop task: A historic footnote", *Consciousness and Cognition*, vol. 12, pp. 347-353, 2003.

[33] 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.

[34] S. Greenfield, *Brain Story,* BBC World Wide Publishing, 2000.

[35] J. G. Taylor and N. Fragopanagos, Modeling the Interaction of Attention and Emotion, *Brain Inspired Cognitive Systems,* Univ. of Stirling, Scotland, UK, August 2004.

[36] 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.*

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

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

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