PERGAMON

Robotics and Computer Integrated Manufacturing 19 (2003) 555–565

Robotics
and
Computer Integrated
Manufacturing

www.elsevier.com/locate/rcim

Towards a human–robot symbiotic system

Kazuhiko Kawamura*, Tamara E. Rogers, Kimberly A. Hambuchen, Duygun Erol

Center for Intelligent Systems, Vanderbilt University, Nashville, TN 37235-0131, USA

Received 8 October 2002; received in revised form 26 April 2003; accepted 23 May 2003

Abstract

Partnership between a person and a robot could be simplified if the robot were intelligent enough to understand human intentions and perform accordingly. During the last decade, we have been developing such an intelligent robot called ISAC. Originally, ISAC was designed to assist the physically disabled, but gradually became a test bed for more robust human–robot teaming (see http://eecs.vanderbilt.edu/CIS/). In this paper, we will describe a framework for human–robot interaction, a multi-agent based robot control architecture, and short- and long-term memory structures for the robot brain. Two applications will illustrate how ISAC interacts with the human.

Keywords: Humanoid robot; Multi-agent system; Human–robot interaction; Behavior learning

1. Introduction

Robotics field has evolved from industrial robots in 1960s to nontraditional branches such as medical robots and search and rescue robots in 2000s. One area that is gaining popularity among robotic researchers is anthropomorphic robots or humanoid robots [1,2]. Increasing popularity reflects a recent announcement of new journal called the International Journal of Humanoid Robotics (IJHR). The inaugural issue of IJHR is expected in 2004 and the Center for Intelligent Systems (CIS) was asked to contribute an article to this inaugural issue [3]. At the Cognitive Robotics Laboratory (CRL) of Vanderbilt University, we have been developing a humanoid robot called the Intelligent Soft-Arm Control (ISAC) (Fig. 1) since 1995. Originally ISAC was designed to assist the physically disabled [4], but gradually became a general purpose humanoid robot to work with the human as a partner or an assistant at home and factory [5]. We have developed a multi-agent architecture for parallel, distributed robot control [6] based on a unique design philosophy [7] as described in Section 2 of the paper, and a robust human–robot interface [8]. Unlike many humanoid research groups in the world who put more emphasis on human-like motion control and efficient walking pattern generation, our group places emphasis on the cognitive aspects of the humanoid. The research described herein is to report recent progress on developing two agents, the Human Agent and the Self-Agent, plus memory structures that enable ISAC to learn new skills. The Human Agent is the humanoid's internal representation of the human. It includes information about the location, activity, and state of the human, as determined through observations and conversations. The Self-Agent is the humanoid's internal representation of itself. It provides the system with a sense of self-awareness concerning the performance of hardware, behaviors and tasks. Our approach to robot memory structures is through short- and long-term memories called the Sensory EgoSphere (SES) and the Procedural Memory (PM), respectively. The SES is a data structure that encapsulates short-term memory for the humanoid in a time-varying, spatially indexed database interfacing the environment with a geodesic hemisphere [9]. It allows ISAC to maintain a spatially indexed map of relative sensory data in its environment. PM is a data structure that encapsulates both primitive and meta behaviors and forms a basis to learn new behaviors and tasks.

*Corresponding author. Tel.: +1-615-343-0697; fax: +1-615-322-7062.

E-mail address: kawamura@vuse.vanderbilt.edu (K. Kawamura).

0736-5845/$ - see front matter
doi:10.1016/S0736-5845(03)00065-6

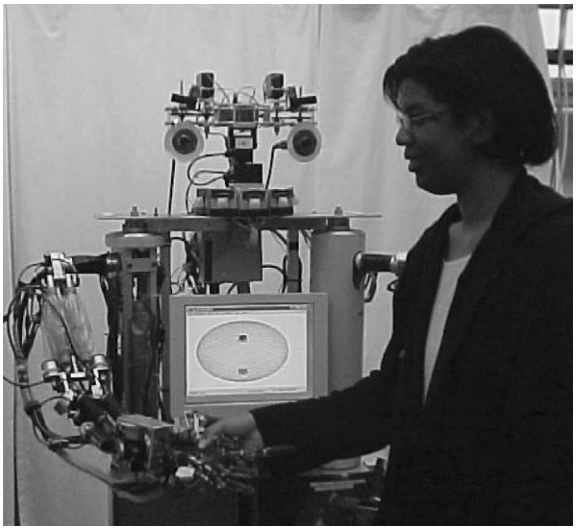

Fig. 1. Vanderbilt's humanoid robot, ISAC.

2. Framework for human–humanoid interaction

Our philosophy for software design for an intelligent machine such as humanoid is to integrate both the human and the humanoid in a unified multi-agent based framework [10]. As such, we have grouped aspects of human–humanoid interaction (HHI) into the following three categories: *Physical, Sensory,* and *Cognitive*. *Physical* aspects are those that pertain to the structure or body of the human and the humanoid. These aspects cover the essential physical features and manipulation capabilities needed for robust interaction. *Sensory* aspects refer to the channels through which the human and the humanoid gain information about each other and the world. *Cognitive* aspects describe those concerned with the internal workings of the system. For the humans these include the mind and affective state; for the humanoid these include the reasoning and abilities to communicate its intention. It is difficult to determine cognitive aspects of humans consistently. Therefore our application is limited to cases where both the human and the humanoid intend to achieve a common goal. We are also interested in giving the humanoid its own emotional or affective module to make HHI more socially appealing [11].

ISAC has been equipped with sensors such as cameras, microphones, and infrared sensors for capturing communication modes (see appendix). We use Microsoft Speech engines for detecting human speech, and have implemented a sound-localization system [12]. An infrared motion detector provides ISAC with a means of sensing human presence. A face detector returns the location of a face and the level of confidence for the detection. A simple finger-point detector locates a person's fingertip.

Likewise, we are developing techniques for ISAC to give feedback to people. ISAC can exploit the sensory modalities that people are accustomed to using, such as seeing and hearing. In addition to speaking to people, ISAC can physically manipulate its body and arms. For example, we are developing display behaviors, such as gesturing that ISAC can use to communicate its intention. We are also developing the use of a visual display of the SES that can be projected on ISAC's monitor located in the middle of its' body (see Fig. 1). The interface is based on multi-agent based architecture [13,14] called the intelligent machine architecture (IMA) developed at Vanderbilt [6]. Fig. 2 illustrates the overall IMA agent structure with the short and long term

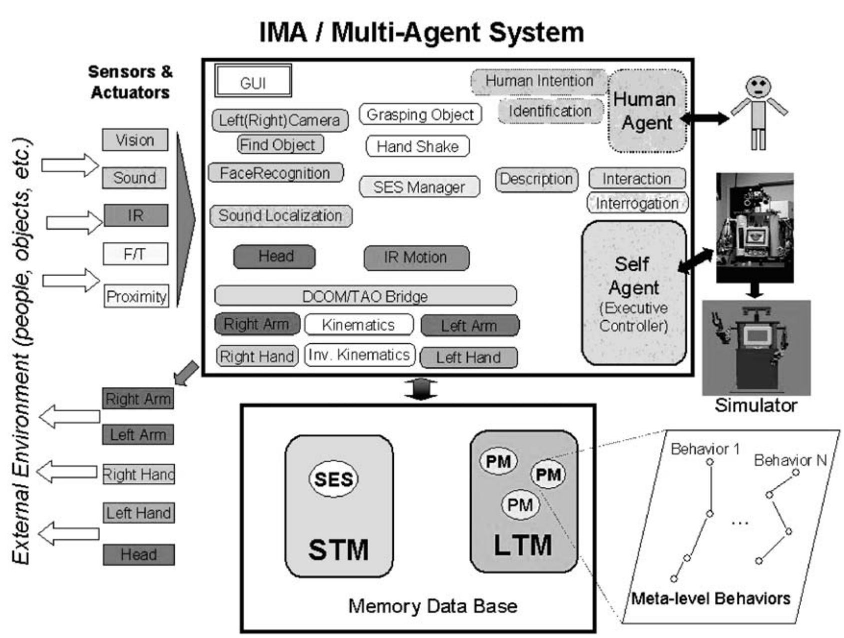

Fig. 2. IMA agents and memory structures.

memory structures. In the sequel, two cognitive agents that form the nucleus of human-ISAC interactions will be described.

2.1. The human agent

The Human Agent [15] is a virtual agent, realized as a collection of IMA agents, which serves as an internal *active representation* of people in the robot's environment. As an *active representation* [16], it is able to detect, represent and monitor people. It also facilitates human-robot interaction by determining appropriate interaction behaviors. As shown in Fig. 3, five IMA agents perform the core functions of the Human Agent. These agents are grouped as two compound IMA agents [6], the Monitoring Agent and the Interaction Agent, which address the two main roles of the Human Agent. The Human Agent receives input concerning the human from various atomic agents that detect the aforementioned *physical* aspects, such as a person's face, speech, location, etc. The Human Agent then communicates with various supporting agents which are each responsible for functions such as detecting features of people, interfacing with memory data structures, or reporting human status information to the Self Agent.

2.1.1. Monitoring agent

This agent monitors human features, including physical features and emotional or affective state. To reflect that people typically interact through various modalities, our approach integrates of several technologies such as speech recognition and speaker identification, face detection and recognition, sound localization, and motion detection, each of which contribute to an awareness of humans and their actions in the environment. The Monitoring Agent includes three atomic IMA agents, the Observer, Identification, and Human Affect Agents. Certain atomic agents, categorized as Human Detection Agents (HDAs) in Fig. 3, each detects a feature of the human, such as the location of a face, a voice, etc., and report this to the Observer Agent. Performance data on various HDAs are shown in Table 1. Two of the HDAs are the Face Detection Agent and the Sound Localization Agent. Face detection is template-based and returns location and confidence of detection. The Sound Localization Agent determines the direction of the source based on the relationship between the energy of the two stereo channels. In turn, the Observer Agent integrates this information to detect and track the human during the interaction and is currently able to integrate data to represent that there are 0, 1, or 2 people in the environment.

The Identification Agent identifies the human in the environment and is used to personalize interactions. It receives information from Human Identification Agents, including the Speaker Identification and Face Recognition agents. Each of these agents employs forced-choice classifiers to match inputs to corresponding data in the library of known people. For speaker identification, we have implemented a simple speaker identification routine using a forced-choice classifier that operates on a library of known speakers stored in the human database. The face recognition algorithm utilizes embedded hidden-Markov models and the Intel Open-Source Computer Vision Library [17]. Location and identity information is posted on the SES, a short-term sensory-event memory of ISAC.

The Human Affect Agent, in a similar design paradigm to the above Observer and Identification Agents, will receive inputs from various agents, called Affect Estimation Agents, which each represent a feature of human affect. Currently, this agent receives input from a Word Valence Agent, which monitors the human's speech for words known to contain positive or negative expression. The goal of this agent is to provide ISAC with knowledge of the person's affective state to ultimately influence the interaction method.

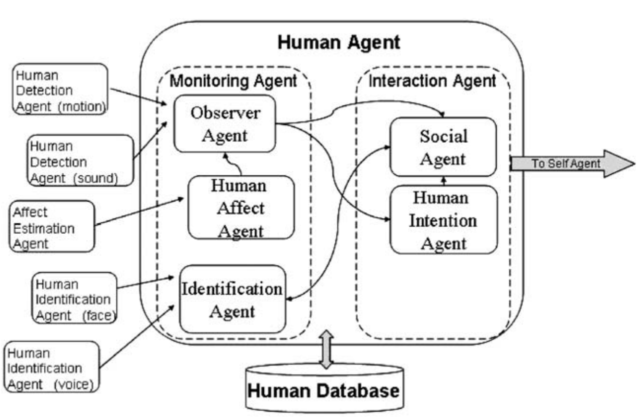

Fig. 3. The Human Agent and supporting atomic agents.

2.1.2. Interaction agent

While the Monitoring Agent performs passive roles of observing people, the Interaction Agent may be

Table 1
Human detection agents

Technology	Accuracy
Sound localization	89% labeling source 6 out of region
IR motion	Reports 72% of time person is there
Human face (detection)	Finds present faces 60%
Human face (recognition)	Correct identity 77%
Speaker ID	Collect further data
Speech recognition	Using Microsoft Speech SDK 5.1
Word valence	(depends on speech recognition—no real measure of accuracy)

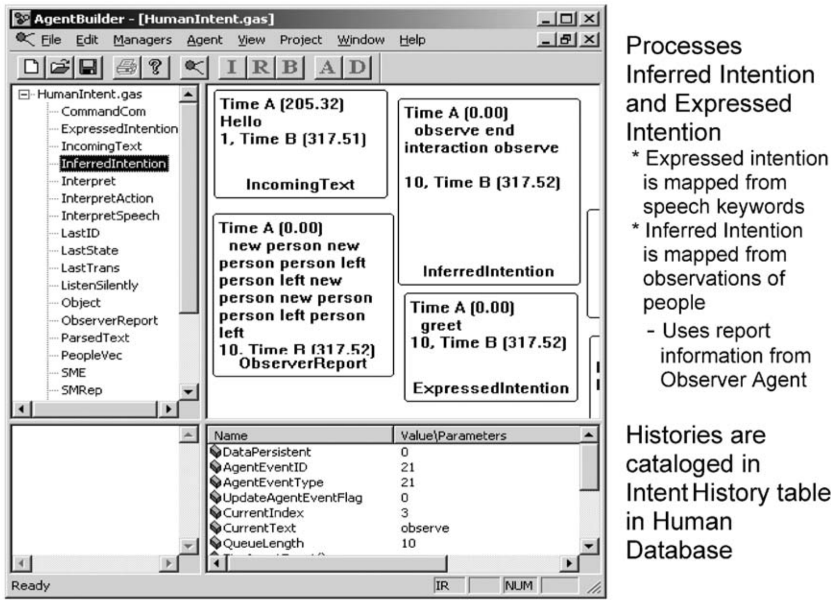

Fig. 4. The Human Intention Agent, shown in an AgentBuilder window.

considered as handling the more pro-active functions of the interaction. Two roles are: (1) handling communication via an Interaction Agent and (2) modeling the interaction with the human via the Social Agent.

The Human Intention Agent, shown in Fig. 4 in an AgentBuilder window, handles verbal and non-verbal communication between the humanoid and the human. The AgentBuilder program is a development tool for the IMA. It provides the interface for the design, testing, and execution of IMA agents. It processes two types of intention from people, namely *expressed* or *inferred*. The Human Agent maps speech directed toward the robot into an expressed intention based on mapping of keywords. If the person greets or requests a task of the robot, this is considered to be an *expressed* intention. Other intentions, however, represent what the robot can infer based on the actions of the human and are labeled *inferred* intentions. For example when a person leaves the room ISAC assumes that the person no longer intends to interact and, therefore, can reset its expectations. An initial suite of human intentions include the intent to communicate (which is used to eliminate speech or sounds with no communicative intent), intent to interact with ISAC, intent for ISAC to perform a task and intent to end interaction. Based on observations of interactions between ISAC and various people, we have included and a special case intent to observe.

The Social Agent contains a rule set for social interaction which enables the robot to interact with people as a function of the overall context and, therefore, more naturally. The method of action selection is based on a set of social rules developed for ISAC to interact with people. The rule base is a production system and operates on features such as the

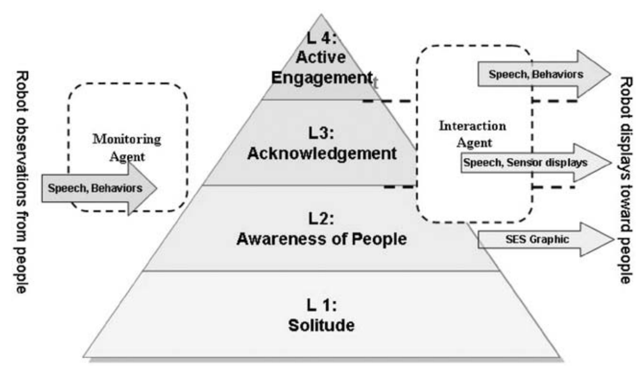

Fig. 5. Levels of interaction within the Human Agent.

level of interaction, current human intention, etc. and provides suggestions of appropriate behaviors to the Self-Agent for consideration. The Social Agent constantly monitors the state of certain variables that represent external information, such as the number of people, current person, and new intentions. The Self-Agent then interprets this suggestion in the context of its own current state, e.g. current intention, status, tasks, etc.

Fig. 5 shows the model of the levels of interaction engagement, represented by a numerical value, that we have developed as the basis for modeling social interaction [15]. These levels progress the robot from a state of no interaction to an ultimate goal of completing a task with (or for) a person. Level 1, Solitude, corresponds to when ISAC does not detect anyone in the environment. In this situation, ISAC may choose actions to actively attract people with whom to interact. Level 2, Awareness of People, corresponds to a stage, often short time, when ISAC is aware of people around it, and has not interacted with them. Level 3, Acknowledgement, is the phase when the robot actively acknowledges the presence of a person. This is performed if the person is approaching ISAC for the first time or if the person is interrupting an ongoing interaction. Level 4, Active Engagement, represents that stage of active interaction.

2.1.3. Human database

The Human Database is a repository of information about people with whom the robot has previously interacted. It provides two services. The first is to store the personal data that the robot uses to recognize and to identify people, images, and voices. Secondly, the Human Database stores data on previous interactions—time, type and individual preferences of interactions or activities. To represent individuals, the database keeps an identification-indexed record of people the robot has met (see Table 2). As the robot learns about a person (presently performed offline), the database contains an indicator that the person has been added to the face and/or voice recognition libraries. These libraries will store the actual processed exemplars for recognition.

The second database table stores task-related information about the interactions that the robot has had with people called representative data (see Table 3). This data is stored by time and is a log of what interactions ISAC has had. This table is the source for determining an individual's most recent or most frequent tasks. It may also be used to determine the robot's most recent or frequent tasks across all people or on a given day. This data will be used to determine personal trends of the current individual and personalize the data.

Table 2
Example of known people

Index	Name	First meeting	Last meeting	Face learned	Voice learned
1	Tamara	2001-12-31	2002-06-18 11:33:34	Yes	Yes
2	Kim	2001-12-31	2002-04-26 19:47:23	Yes	Yes
3	Xinyu	2001-12-31	2002-06-20 10:15:32	Yes	Yes

Table 3
Example of intention histories database table

Index	Name	Intent	Time
1	Tamara	ColorGame	2002-06-18 11:33:34
2	Xinyu	Recognize	2002-06-20 09:35:54
3	Xinyu	HandShake	2002-06-20 09:37:23
4	Xinyu	ColorGame	2002-06-20 09:39:43

2.2. The Self-Agent

The Self-Agent is a cognitive agent responsible for monitoring sensor signals, agent communications, and high-level (i.e. cognitive) decision-making aspects of ISAC (Fig. 6). Sensory signal monitoring currently involves component-level fault monitoring and diagnosis using Kalman filter.

The Self-Agent then integrates failure information from sensors and maintains information about the task-level status of the humanoid. Cognitive aspect covered by the Self-Agent includes recognition of the human's intention it receives from the Human Agent and to select appropriate actions by activating meta behaviors within the long-term memory (LTM). Fig. 7 shows the structure of the Self-Agent and supporting atomic agents.

The Intention Agent will determine the intended action of the humanoid based on the intention of the human, the state of the humanoid, and whether this action would conflict with the humanoid's current activities. If there is no conflict, the action can be performed, otherwise, the Intention Agent must resolve the conflict. Conflict resolution is handled based on the relative priority of the conflicting actions. Because ISAC currently can only perform a limited number of actions, we use a table ranking possible actions by priority. When the current action's priority is higher than the new intended actions, then the current task will be paused. ISAC will explain to the person why the new action cannot be performed and will resume current activity. If the new action's priority is higher, then the current action will be halted and the new action will begin.

The Central Executive Controller (CEC), under construction, coordinates the various PM structures stored in the LTM that encapsulate ISAC's behaviors. These behaviors are used in a plan constructed by the CEC to perform a task. The goal of the generated plan is determined according to the input from the Intention Agent, then the constructed plan is put into action. The execution of the plan is monitored by the Self-Agent and in the case of an abnormal execution, the Self-Agent is responsible for solving the impasse by generating a novel plan.

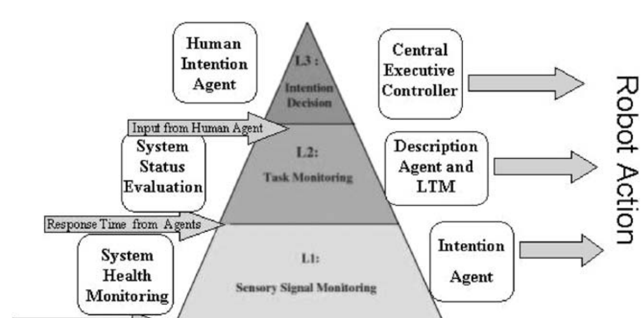

Fig. 6. Levels of monitoring and action in the Self-Agent.

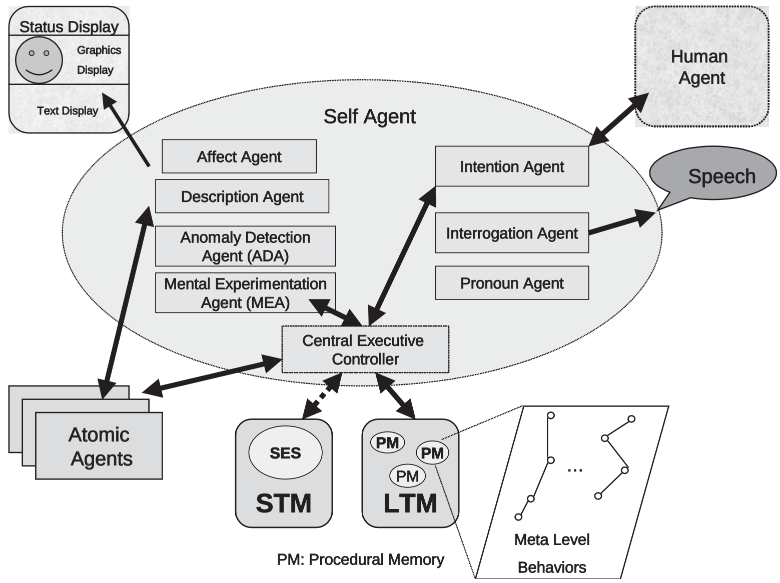

Fig. 7. The Self-Agent and supporting Atomic Agents.

3. Robot memory data structures

To monitor the vast amounts of data and procedures a robot acquires when interacting with people, a robot should maintain both short- and long-term memory capacities. These memory structures should store and manipulate sensory, declarative and procedural information. Two data structures have been developed for ISAC to address these issues: the SES provides ISAC with a short-term memory store for sensory events while Procedural Memories provide ISAC with a long-term memory for performing behaviors.

3.1. STM: SES

The SES is a data structure that serves as a short-term memory for a robot [9]. The SES was inspired by the egosphere as defined by Albus [18] Albus and Meystel [19]. The original egosphere was conceived as a topological sphere surrounding an entity onto which external or internal events are projected. Albus proposed multiple egospheres for different components of a system, i.e. head egosphere, camera egosphere, body egosphere. The objective of the SES is to store sensory information, both exteroceptic and proprioceptic that the robot detects. The SES is structured as a geodesic sphere that is indexed by azimuth and elevation angles and is centered at a robot's origin. With ISAC, the head's pan-tilt origin serves as the center of the SES. The geodesic dome provides an interface for projection of data while a database provides storage for a description of projected data. Sensory processing agents in the system send their outputs to the SES for short-term storage. The SES then projects these outputs onto its geodesic interface. The projection vector is a unit vector from the center of the sphere in the direction of the robot's pan and tilt angles. The information and data that was sent by the sensory processor is then entered into the database with a timestamp reference.

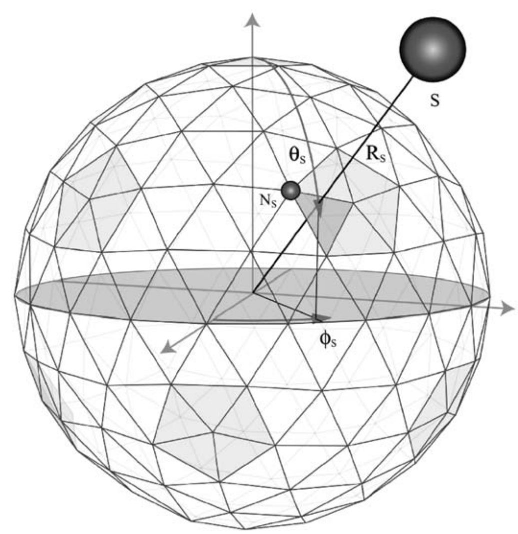

Fig. 8. Projection of objects onto the SES.

Fig. 8 shows the projection of an object onto the SES. In this view, ISAC's head is located in the center of the sphere. The projection vector occurs at a given azimuth, elevation from the center of the sphere which correlates to a specific pan, tilt location of the head. Once this object is projected onto the sphere, a distance measure is performed to find the vertex on the sphere closest to the projection spot. This vertex, or node, becomes the registration node for the specific object. The SES enters the object, any information reported about the object and a timestamp into the database at the location of its registration node. This process occurs each instance that a sensory processing agent sends its output to the SES.

The SES can also retrieve information stored in its database. An agent can request retrieval of data from the SES in one of three methods: by data name, data type of location of data. For the first two methods, the SES simply queries the database using the name or type of the data. To retrieve data using a location, the requesting agent must supply the SES with a specific pan, tilt angle pair and a search neighborhood. The search neighborhood specifies how many nodes from the center node to retrieve data. The center node is defined as the closes node to the given pan, tilt angle pair. The SES calculates which nodes are included in the neighborhood and then queries for data from these nodes. Fig. 9 shows ISAC, the SES and the registered location of events ISAC has detected.

Since the SES is a short-term memory component, old data is continuously removed from the sphere. The timestamps of each registered data are checked against pre-defined time limits. If any of the timestamps exceed the limits, the data is removed from its registration node.

3.1.1. SES graphical interface

A graphical display of the SES is an interface for assisting the human to understand what ISAC is sensing in the environment. The prototype display (Fig. 10) shows ISAC's view of the SES. Each time an object is posted to the SES or old records are purged from the database, the SES sends out all contents in ISAC's short-term memory to the display agent. The display shows icons of the types of data posted and labels the icons with the data names, thus the human sees what ISAC senses. This view provides the person with insight into malfunctions in the system, such as when a sensory event occurred, but was not detected by ISAC.

3.2. LTM: PM

LTM is a data structure which contains the primitive and meta-level behaviors that will be combined later to perform the desired tasks. We call a LTM unit member a Procedural Memory (PM). The PM encapsulates primitive and meta-level behaviors (Fig. 11). These behaviors are derived using the spatio-temporal Isomap method proposed by Jenkins and Matarić [20]. A short description of how it was used to generate LTM units is shown.

Motion data are collected from the teleoperation of ISAC. The motion streams collected are then segmented into a set of motion segments as shown in Fig. 11. The central idea in the derivation of behaviors from motion segments is to discover spatio-temporal structure in a motion stream. This structure can be estimated by

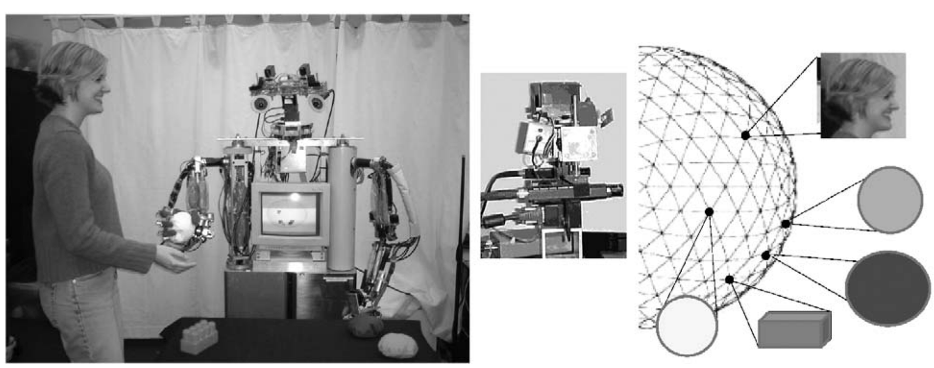

Fig. 9. Projection of objects and events onto the SES.

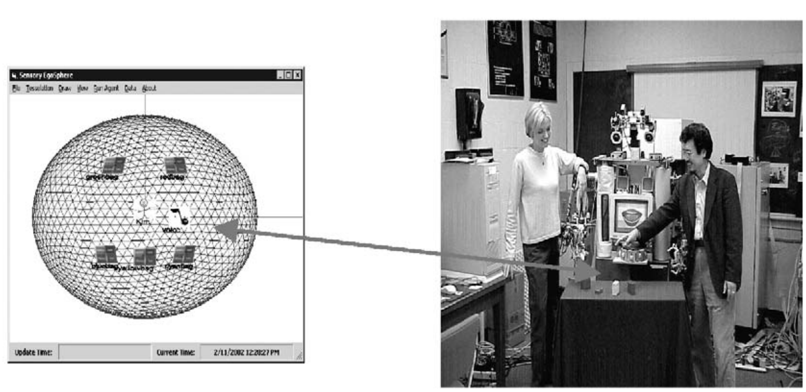

Fig. 10. Graphical display of the SES.

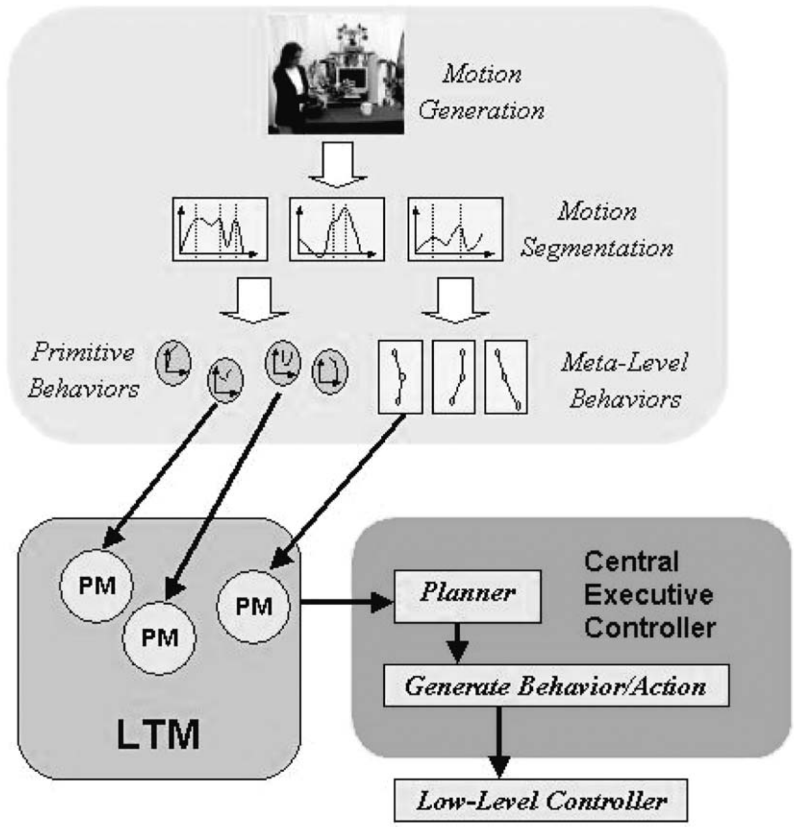

Fig. 11. Behavior representation in LTM.

extending a nonlinear dimension reduction method called Isomap [21] to handle motion data with spatial and temporal dependencies. Isomap method was originally developed to handle a large number of spatially distributed data. It was extended by Jenkins and Mataric to generate spatial and temporal motions generated by human motions. Spatio-temporal Isomap dimension reduction, clustering and interpolation methods are applied to the motion segments to produce primitive behaviors (Fig. 11). Meta-level behaviors are formed by further application of the spatio-temporal Isomap method and linking component primitives with transition probabilities [22].

We are currently developing an agent called the central executive controller (CEC) which generates a task-level behavior or action through a combination of planning mechanism, goals, intentions and beliefs.

4. Applications

4.1. Demonstrations

4.1.1. Demonstration in human-guided object learning through shared attention

This demonstration through finger pointing utilizes the HHI framework of Fig. 2 to allow people to direct ISAC's attention to objects in its workspace. In the current demonstration, represented in Fig. 10, ISAC is directed by a human to look at several objects (assorted color blocks) on a table. When ISAC is told to look at the green block, ISAC looks for the pointed finger. ISAC then takes the position of the green block and registers the location and name onto the SES. ISAC is directed to look at a red block and repeats the previous actions. After the blocks are registered onto the SES, ISAC returns to its initial position (looking straight ahead). Then ISAC is told to look at one of the previously taught objects. ISAC retrieves the object named from the SES and saccades to the location given in the SES.

The application involves several levels of the framework for directing attention to known and unknown objects. Two aspects of this interaction, learning and recalling object locations are shown schematically in Fig. 12. To direct ISAC's attention to an unknown object (Fig. 12a), the application begins with speech from the human directing the robot's attention to an object. The Human Agent activates the Human Finger Agent and parses the name of the object. The Human Finger Agent finds a pointed finger to fixate on the object. At this point, the Human Agent sends the object name and location to the SES Manager, which registers it on the SES. To direct ISAC's attention to a known object that

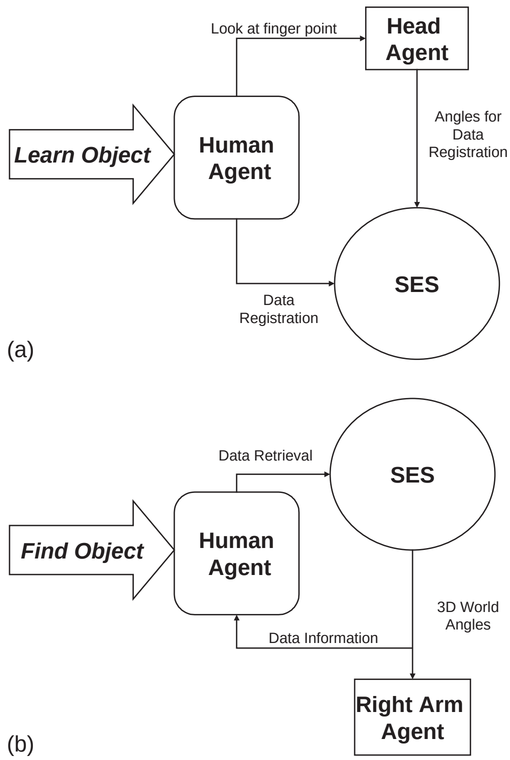

Fig. 12. Schematic for (a) learn object and (b) find object.

is registered on the SES, the human tells ISAC to find the desired object. The Human Agent translates this speech command into text and sends the object name to the SES Manager to initiate a search of the SES using the name. When the object is retrieved, the SES Manager returns the location of the object to the Human Agent. In the case of a person requesting ISAC to recall the location of an object, the flow, depicted in Fig. 12b, is as follows. The person's request is parsed by the Human Agent into text describing the object of interest. This information is forwarded to the SES to retrieve the object's location. When this information is returned, the world coordinates of the object are sent to the Right Arm Agent and the robot performs a pointing gesture to point to the requested object.

4.1.2. Demonstration in situation-based acknowledgement

In this demo, ISAC processes the intentions of the human, resolves them with its own intentions and abilities, and communicates to the person if there is a problem with the request. The scenario begins as a person approaches ISAC and gains its attention. The Human Agent determines that the person has an intention to interact with ISAC. If ISAC is unoccupied at the time, ISAC begins its interaction behaviors by turning toward the person and initiating a greeting and identification sequence. Once interaction is established, ISAC begins a social dialog. After greeting, ISAC may respond to a person's task request (an intention for ISAC to do something) if it is within ISAC's abilities. If a second person approaches ISAC and attempts to gain its attention, the Human Agent will notify the Self-Agent that there is a new person with a pending intention. The Self-Agent then must resolve the current human intention with its own current intention. If the second human intention is not of sufficient priority to override ISAC's current task, then ISAC will then pause its current interaction, turn to the interrupter, and apologize for being busy (Fig. 13b). ISAC can then return its attention to the first person and resume its previous interaction (Fig. 13a). There may also be a situation where the request of the interrupting person actually has higher priority than the task ISAC is currently performing. In this case the Self-Agent determines to switch to the new task after giving an explanation to the current person.

4.2. Future enhancement

To demonstrate the role of LTM in conjunction with STM, we are planning to apply the methodology shown

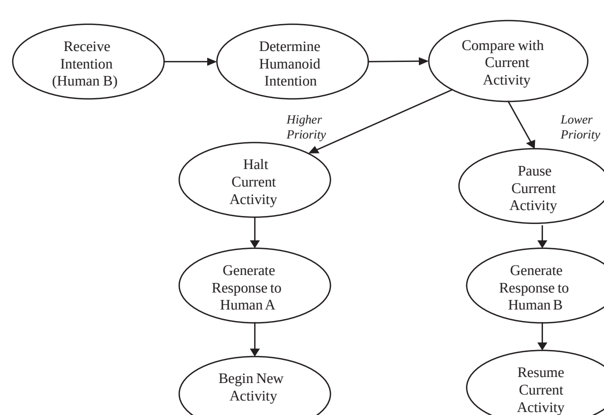

Fig. 13. (a) ISAC responding to an interruption; (b) role of intention processing during an interruption.

in Fig. 11 to the HandShaking Demo [15]. A set of meta-level handshaking behaviors will allow ISAC to adapt its behavior when a new person comes and greets ISAC from unexpected directions.

5. Conclusion

Realization of general-purpose humanlike robots with adult-level intelligence continues to be the goal of many robotic researchers. We have already seen major advances in the last 5 years in humanoid robotics research and expect this progress to accelerate. In particular, we may see a major breakthrough in terms of new behavior and task learning within the next several years. We expect ISAC to adapt its behaviors and learn new tasks based on its self-reflection mechanism and external interactions with humans and the environment in limited domains [23]. A growing number of robotics researchers believe that intelligence is an emerging property of an autonomous agent such as humans and robots and that behavior learning requires the coupling of mind, body and the environment. It is our goal to make human–robot interaction less hard-coded and more adaptive and socially acceptable through cognitive robots.

Acknowledgements

This work has been partially funded through a DARPA grant (Grant ♯DASG60-01-1-0001) and a NASA GSRP (Grant ♯NGT9-52-JSC). The authors would like to thank colleagues and graduate students in the Center for Intelligent Systems at Vanderbilt University for their contributions.

References

[1] Humanoid Robotics, IEEE Intelligent Systems and their Applications, Special Issue 15:4, July August 2000.
[2] Fukuda T, et al. , How far away is artificial man? IEEE Robotics Automat Mag 2001;7(1):66–73.
[3] Ming X. Private communication, April 15, 2003.
[4] Kawamura K, Bagchi S, Iskarous M, Bishay M. Intelligent robotic systems in service of the disabled. IEEE Trans Rehabil Eng 1995;3(1):14–21.
[5] Kawamura K, Wilkes DM, Pack T, Bishay M, Barile J. Humanoids: future robots for home and factory. Proceedings of the First International Symposium on Humanoid Robots, Tokyo, Japan, 1996. p. 53–62.
[6] Pack RT, Wilkes DM, Kawamura K. A software architecture for integrated service robot development. Proceedings of the IEEE Systems, Man and Cybernetics (IEEE-SMC), 1997. p. 3774–9.
[7] Kawamura K, Pack RT, Bishay M, Iskarous M. Design philosophy for service robots. Robotics Autonomous Syst 1996; 18:109–16.
[8] Kawamura K, Alford A, Hambuchen K, Wilkes M. Towards a unified framework for human-humanoid interaction. Proceedings of the First IEEE-RAS International Conference on Humanoid Robots (Humanoids'00). Cambridge, MA: MIT; 2000.
[9] Peters II RA, Hambuchen KE, Kawamura K, Wilkes DM. The Sensory EgoSphere as a short-term memory for humanoids. Proceedings of the Second IEEE-RAS International Conference on Humanoid Robots (Humanoids2001), 2001. p. 451–9.
[10] Kawamura K, Peters II RA, Wilkes MW, Alford WA, Rogers TE. ISAC: foundations in human–humanoid interaction. IEEE Intell Syst Appl 15:4, July August 2000. p. 38–45.
[11] Breazeal C. Designing sociable robots. Cambridge, MA: MIT Press; 2002.
[12] Sekman AS, Wilkes M, Kawamura K. An application of passive human–robot interaction: human tracking-based on attention distraction. IEEE Trans Syst Man Cybern: Part A. Syst Humans 2002;32(2):248–59.
[13] Minsky M. The society of mind. New York: Simon and Schuster; 1986.
[14] Ferber J. Multi-agent systems: an introduction to distributed artificial intelligence. Harlow England: Addison-Wesley; 1999.
[15] Kawamura K, Rogers TE, Ao X. Development of a human agent for a multi-agent based human–robot interaction. AAMAS 2002. p. 1379–86.
[16] Bajcsy R. Active perception. Proc IEEE 1988;76:996–1005.

Appendix A. ISAC system information

Interaction competencies for a personal robot	Hardware	Software
Observational		
Presence of person: infrared (IR)	Passive IR motion detector array	Digital I/O
Sound: event localization	Condenser Microphones	Matlab
Speech: detection/recognition	Handheld	Microsoft Speech Recognition engine 4.0
Vision: face and gesture	Sony Color CCD Cameras	Visual C++ routines, some with intel Libraries
Demonstrative/responsive		
Speech: synthesis	PC Speakers	AT&T Natural Voices Engine
Motor behaviors: head	Directed Perception PTU-46-70	IMA Wrapper for serial port
Motor behaviors: arms	Pneumatic muscles	Visual C++ routines and control

[17] Nefian A, Hayes M. Face recognition using an embedded HMM. Proceedings of the IEEE Conference on Audio and Video-Based Biometric Person Authentication, 1999. p. 19–24.
[18] Albus JS. Outline for a theory of intelligence. IEEE Trans Syst Man Cybern 1991;21(3):473–509.
[19] Albus JS, Meystel AM. Engineering of mind: an introduction to the science of intelligent systems. New York: Wiley; 2001.
[20] Jenkins OC, Matarić MJ. Automated derivation of behavior vocabularies for autonomous humanoid motion, Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems, Melbourne, Australia, 2003.
[21] Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. Science 2000;290(5500):2319–23.
[22] Erol D, Park J, Kawamura K, Turkay E, Jenkins OC, Matarić MJ. Motion generation for humanoid robots with automatically derived behaviors. Proceedings of the IEEE Systems, Man and Cybernetics, Washington, DC, October 2003.
[23] Kawamura K, Noelle DC, Hambuchen KA, Rogers TE, Turkay E. A multi-agent approach to self-reflection for cognitive robots. Proceedings of the 11th International Conference on Advanced Robotics, Coimbra, Portugal, June 30–July 3, 2003.