The Human Agent: a work in progress toward human-humanoid interaction

Tamara Rogers and Mitch Wilkes
Vanderbilt University
Nashville, TN 37235

Abstract

At the Intelligent Robotics Laboratory of the Center for Intelligent Systems at Vanderbilt University, we have been developing a humanoid system called ISAC, (Intelligent Soft-Arm Control), over the past several years. As people work with a humanoid system, the interaction should be natural and robust. Our framework for human-humanoid interaction (HHI) considers various aspects of HHI and this paper will present information about the development of one key element of this system, the Human Agent.

1 Introduction

At the Intelligent Robotics Laboratory of the Center for Intelligent Systems at Vanderbilt University, we have been developing a humanoid system called ISAC, (Intelligent Soft-Arm Control), over the past several years. ISAC (see Figure 1), designed and built in the IRL, was originally developed for a robotic aid system for the physically disabled. In service robotics, interaction with humans is essential to the robot's ability to carry out its function. Over the years, ISAC has grown into a testbed for research in human-humanoid interaction. Although the system continues to address the issue of aiding humans, our recent research emphasis has been on general problems of human-humanoid interactions.

Implied in the use of the term humanoid is the belief that the robot not only looks somewhat human but also behaves in a manner that is perceived as intelligent. Well-designed interaction imparts a sense of believability to the robot. People seem to perceive robots as more believable when the interaction is in some sense similar to the way humans naturally communicate. For example, the robot should be able to autonomously detect the human and what he is doing. Human-Humanoid Interaction (HHI) is described by a multi-agent representation inside the robot.

This paper is organized as follows. Section 2 briefly introduces our humanoid system. Section 3 describes our human-humanoid interaction research directions and the framework that is our approach toward achieving these goals, including several key software elements. One of these elements, the Human Agent, which is the focus of this work, is presented in section 4. Section 5 presents two scenarios that demonstrated the feasibility of this work. Section 6 presents conclusions and future work directives.

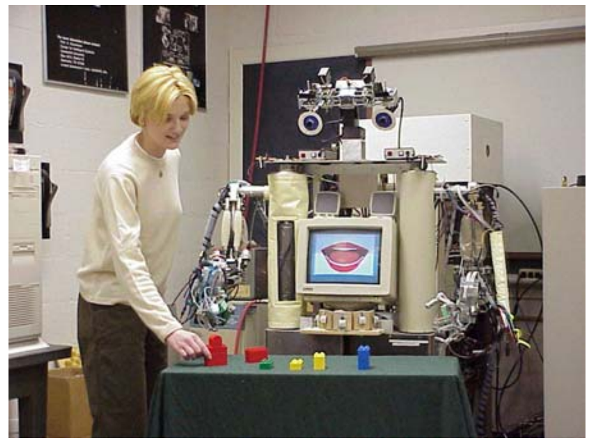

Figure 1: ISAC, our humanoid system

2 Humanoid System

Our humanoid system, ISAC, was designed using the philosophy of embodied intelligence [2]. That is to say that in order to achieve human-like interaction with people, the robot should have a humanoid shape. ISAC has two 6-degree-of-freedom arms actuated by McKibben artificial muscles. ISAC's end effectors, or hands, the PneuHand I and PneuHand II, are anthropomorphic and were built in-house. PneuHand I is pneumatically actuated while the PneuHand II combines electric and pneumatic actuation. The arm-hand systems have 6-axis force-torque sensors at each wrist, proximity sensors on the palms, and rudimentary touch sensors on each finger. ISAC also employs a stereo vision system with color digital cameras, an active camera head and eyebrows for emotion feedback via facial expression. Microphones provide the capabilities for speech processing and sound source localization. A passive infrared sensor array for warm-body motion detection completes the sensory suite.

3 HHI Goals and Framework

An overriding goal for our humanoid system is to enable natural unfettered interactions with people. That is to say, the human should not have to wear special device or otherwise be physically connected to the system. A long-term HHI goal for our system is to provide tools to allow humans to interface with a humanoid as a member of a team. Therefore, subtasks of this goal include enabling the system to detect, monitor, identify and socialize with humans in the system's environment. This includes determining how many people, if any, are present in the environment. For persons in the environment, the system should also determine where they are, who they are, what they are doing and how they are doing (i.e., what is their emotional state). The realization of these functions requires the implementation and integration of such technologies as human detection and localization, person identification, speech and gesture recognition, and affective computing. Our short-term HHI goals consist of developing these abilities for the case of a single person in the humanoid's environment. The system should be able to detect the person's presence, determine either the person's identity or that this is a new interaction, communicate with the person concerning tasks, and dynamically decide some measure of the human's state of satisfaction or frustration with the robot's behavior. Intermediate goals in this work expand the single person case to that of two people in the environment. The increased complexity lies in the ability of the system to distinguish dynamically between the two people as they move about the environment. The two-person case provides a situation for the system behavior to physically demonstrate that attention is being switched from one person to the other.

3.1 HHI Framework

Our humanoid system is being developed within a multi-agent robot software architecture. The software runs on a Windows NT network of PCs and is developed using the Intelligent Machine Architecture (IMA) developed in our lab. The IMA taxonomy provides for *atomic agents* that are built out of IMA *components*, which are DCOM objects, and have one or more threads of execution and are independent, autonomous entities. Atomic agents are usually categorized by their function, for example, hardware/resource agents, behavior agents or sequencer agents. *Compound agents* are collections of atomic agents that communicate and interact and are used to achieve useful results. IMA agents can provide the representation and action for elements of the humanoid, the environment or the knowledge base.

ASPECTS	***Human***	***Humanoid***
Physical	Face	Structure
	Hand	Workspace
Sensory	Audio	Audio
	Vision	Vision
	Others	Tactile
		Infrared
Cognitive	Emotion	Affect
	Logic	DBAM
		SES

Figure 2: Aspects of HHI for developing an interface

The humanoid system is being developed within a framework developed to promote useful HHI, see Figure 2. This framework addresses the desire to develop a humanoid that detects interesting aspects of humans. These include the physical, sensory and cognitive aspects of interaction. The humanoid should be able to detect these aspects of humans and present its own versions as well. Physical features of humans that are commonly used when people interact include faces and hands, for example. Therefore, ISAC should detect these during interactions, as well as use its analogous physical features to communicate back to the human. Hence, when meeting a person, ISAC makes eye contact and offers its hand for a handshake. The sensory aspects include vision, audition, and touch. The humanoid can express or detect these via technologies such as speech synthesis, speech recognition, and color vision. The cognitive aspects are more difficult to implement. Detecting a human's emotion and modeling affect in humanoids are both open research issues.

To address our HHI goals within this framework, we propose high-level elements, such as the Self Agent, the Human Agent, the Sensory EgoSphere (SES), and the Database Associative Memory (DBAM). These key parts of the system are shown in Figure 3.

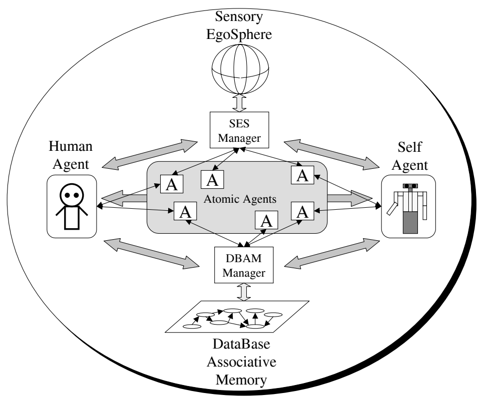

Figure 3: Key elements of the HHI Framework

To facilitate human-like interaction, an "inside ISAC/outside ISAC" distinction for social interaction is provided by two high-level compound agents. The Self Agent provides the system with a sense of what is happening inside the humanoid, including goals and status. The Human Agent provides the system a way of representing humans in the environment and their activities. It coordinates the other agents that are used for interacting with humans. The Sensory EgoSphere is a repository for sensory events in the environment and functions as a short-term spatially-oriented memory. Events in this database are indexed by the direction of occurrence from the robot's center (ego). The Database Associative Memory is under development and will responsible making associations between conditions and actions that the humanoid has stored in this memory. It is expected to serve as a longer-term action-based knowledge memory. The remainder of paper will focus on the Human Agent. Its job is to detect, monitor and interact with humans that are in its environment. This agent allows the human to interact in familiar ways and to interface with the system's awareness of itself, in the Self Agent.

4 The Human Agent

4.1 Decomposition

The Human Agent can be considered on two different axes of decomposition. The first is the logical decomposition based on function. The second is the implementation decomposition based on the IMA agent structure. The Human Agent performs the functions of detection and identification of humans and monitoring human intentions through its interaction. To provide these functions, the Human Agent is divided into the three following logical pieces that monitor and process sensory data.

- Human Monitoring - detects and monitors the location and features of the humans in the environment.
- Identification - determines the identity, if possible/known, of the humans in the interaction
- Social Interaction - coordinates social skills and behaviors, such as spoken responses, greeting behaviors, etc.

These logical elements of the Human Agent are realized through one or more IMA agents.

Human Monitoring, which detects and tracks humans in the environment, is deliberately multi-faceted. It uses visual face detection routines, a passive infrared sensor array, and a sound source localizer to achieve its goals of locating and tracking humans. The distributed software architecture allows the system to use a combination of these technologies without requiring that they each be functional. This idea is demonstrated in Sekmen, et. al. in this same volume. Data from these technologies is stored on the Sensory EgoSphere. Posting the data to the SES structure provides the system with a repository for location and feature data concerning the human. All agents have access to this data stored on the SES. Human Monitoring uses temporal information to make decisions concerning the age and priority of the human-related data.

Social Interaction is responsible for receiving the input to the system and choosing appropriate social responses. This includes, not only speech or textual communication, but also nonverbal behaviors that humans are accustomed to detecting in interpersonal communication. There are agents that perform the functions of processing speech data and finger pointing gestures to determine the information that the humans attempt to communicate with the system. Textual information is processed in coordination with the Interrogation and Interaction Agents in the Self Agent. In addition, a model for social interaction find cite is being developed to allow the system to choose appropriate social behaviors, such as waving, maintaining or breaking eye contact, and offering a handshake.

Identification is responsible for learning characteristics of humans, storing records and matching them to identify people. Currently, Identification uses short-term memory information about a person, including name, height and clothing color. This provides the system with an indication of whether the human in its attention has changed. Other identification approaches

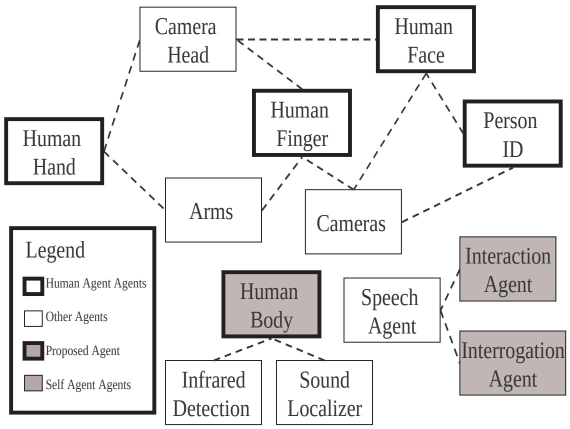

Figure 4: IMA Agents comprising the Human Agent

have been investigated, including a simple voice recognizer that determines a speaker from a library of known speakers.

4.2 Interaction with other agents

The Human Agent is a compound agent in the IMA taxonomy[3]. As a compound agent, it has several constituent atomic agents that are responsible for various functions.

Figure 4 shows a partial breakdown of the Human Agent viewed as its IMA implementation. In the Human Agent there exist atomic agents that are responsible for particular parts of the interaction. For example, in this figure, Human Face is responsible for face detection and tracking. Therefore, it initiates the camera head direction, begins the face finding algorithm and controls the face tracking behavior. When the Camera Head has fixated upon the face, the Human Agent determines the next response in the interaction, depending upon if the goal is to simply locate the human or to maintain eye contact. Hence, Human Face is logically part of Human Detection and Monitoring. Similarly, Human Finger is responsible for locating a human pointed finger. In a manner similar to that of the Human Face, this atomic agent controls camera head location and image processing algorithms for detecting the location pointed by the human.

4.3 Integration of I/O technologies

Our approach is the integration of several technologies that contribute to an awareness of humans in the environment and what they are doing. The nature of the IMA allows the components of the system to be distributed. The ability to detect human presence is given by an array of passive infrared motion sensors, augmented with audio and visual detection and tracking. The audio components include sound-source localization and speech recognition capabilities. The visual components include face detection/tracking and pointing finger detection/localization within the arm's workspace. Simple methods of identification have been implemented and integrated. These are based on the person being introduced to the system, at which time the system then learns the person's name and associates to them their height and the color clothing worn at the time of the meeting. This provides the system with a type of short-term memory for people. Combining the identity and location information will be the key to expanding the interaction to more than one person. Communication with the human is primarily based on speech. Currently our system is in a noisy environment and the benefit of integrating communication modalities is clear. Our current gesture interface is limited to using a pointing finger to supplement the human's speech in certain deictic contexts. Within the multi-agent architecture, the Human Agent is free to communicate with other agents and processing modules. By providing the communication interface between the human and the humanoid, the human may also query the Self Agent concerning the system's internal status.

5 Proof of Concept Examples

5.1 Human Detection

This demonstration is more greatly reported in [6, 5]. The purpose of this demonstration was to verify the ability to use various distributed modes of detecting a human, face detection, sound source localization and passive infrared, to make human detection more robust. Face detection was used as the main modality, with the other two (sound and infrared) used to refocus face detection when appropriate. By *refocus*, we mean to cause ISAC to look at a face in a new direction. The system was allowed to be refocused according to an attention model. In this model, each modality created an time-decaying attention signal. The signals from sound localization and motion detection are inhibited by that of face detection. The model does not allow the system to be distracted until after the face signal has decayed significantly [5].

Two situations were explored. In the first situation, a person was tracked visually. If ISAC's face routine became confused or distracted from the person, the person could use sound or motion to turn the camera head toward the person creating the event. This showed the benefit of using complementary technologies to make human detection more robust. In the second situation, two people were able to vie for ISAC's attention. For example, if one person interacted with the robot, a second person attempted to gain the robot's attention using sound or motion. The second subject's success depended upon the novelty of the first interaction. If the system had just begun its interaction with a person when a distraction occurred, the distraction was unsuccessful. Yet, if ISAC and a person had been interacting for a while, ISAC was able to be distracted more easily by the presence of an additional person.

This work has obvious uses for expanding the system to operate smoothly with multiple people in the environment. It also sets the way for work to allow ISAC to develop some social graces by parameterizing its "distractability" to study effects of a polite vs. rude humanoid system.

5.2 Person Identification

This demonstration was built to demonstrate the ability of the system to autonomously determine information about a person, store it, and retrieve it with certain criteria to determine if it "knows" the person presently before it.

In this scenario, ISAC was introduced to five different people. The system learns each person's name, height and shirt color. The height is determined via the stereo vision system when eye contact is made. The shirt color was also determined using color models and offset from the fixation point. After the system had stored the representations of the people, one person was allowed to interact with ISAC. The system dynamically determined these features for that person and matched them to the database to identify the person. ISAC then monitored the person, continuing to dynamically determine features. Monitoring was tested as the person switched places with another subject. The monitoring capabilities determined that the features had changed significantly to indicate that a new person was now before ISAC. This included that the person's height changed more than a prescribed amount, and or that the clothing changed color. ISAC then responded with the name of the person it presumed to be now in its presence.

5.3 Learning object location through pointing

The purpose of this scenario was to demonstrate the ability of the humanoid to learn, via the Human Agent and the SES, the location of objects that the human is pointing out and to demonstrate recall of them upon demand. This demonstrates the feasibility of a short-term interaction memory that allows the system to learn.

Figure 5: Scene from pointing demonstration

In this scenario, a person pointed to an object and told ISAC to "learn" the object as in Figure 5. At this point the system looked for the person's finger and fixated upon it. This location is used to store the object in the SES. After repeating this procedure for multiple objects, ISAC is asked to find a particular object. The location was retrieved from the SES. ISAC then demonstrated the ability to recall by looking at the object requested. In other work, ISAC has also shown the indicated direction by pointing with an arm.

6 Conclusions and Future Work

The capabilities presented in the above demonstrations show the suitability of the distributed nature of the IMA and the HHI framework, especially the progress of the Human Agent. Future work on the Human Agent includes completing short-term goals and expanding to the two-person case.

One person - short-term goals to be completed include:

- Identification - fully integrate the speaker identification capability to automatically run at the beginning of an interaction. Also, future work will be expanding the personal information determined for people that the humanoid "learns."
- Social Interaction - develop the model of interaction that determines the nature of the human's communication and the appropriate responses for the humanoid.

Multiple people - The foundation for expanding this work to two persons has been demonstrated by the object learning example utilizing the SES, the Person Identification demo and the Human detection scenarios described above. The expansion of identification capabilities will allow the humanoid to detect and remember that there is more than one user in its environment. The work integrating the SES will allow multiple humans to be remembered and recalled individually. Not only this, but the humanoid will be able to alter its response based on what information it has obtained about the individuals in its environment.

Acknowledgments

This work has been partially funded by the DARPA Mobile Autonomous Robotics Systems (MARS) Program.

References

[1] W.A. Alford, T. Rogers, D. M. Wilkes and K. Kawamura, " Multi-Agent System for a Human-Friendly Robot," *IEEE International Conference on Systems, Man and Cybernetics*, pp. 1064-1069, Tokyo, Japan, October 1999.

[2] R.A. Brooks, "A Robust Layered Control System for a Mobile Robot," *IEEE Journal of Robotics and Automation*, 2(1), 1986, pp. 14-23.

[3] K. Kawamura and A. Alford and K. Hambuchen and M. Wilkes, "Towards a Unified Framework for Human-Humanoid Interaction", submitted to *Proceedings of the First IEEE-RAS International Conference on Humanoid Robots*, September 2000.

[4] R.T. Pack, D.M. Wilkes, and K. Kawamura, "A software architecture for integrated service robot development," *IEEE International Conference on Systems, Man and Cybernetics*," pp. 3774-3779, Orlando, FL, 1997.

[5] A. Sekmen, "Toward A Human-Robot Interaction Methodolgy," *Ph. D. Dissertation*, Vanderbilt University, August 2000.

[6] A. Sekmen, A. Alford, T. Rogers, and M. Wilkes, "Human Tracking Based on Attention Distraction," submitted to*IEEE International Conference on Systems, Man and Cybernetics*," Nashville, TN, October 2000.