Modeling Human-Robot Interaction for Intelligent Mobile Robotics

Tamara E. Rogers, Jian Peng, and Saleh Zein-Sabatto

College of Engineering, Technology, and Computer Science
Tennessee State University
Nashville, TN 37209, USA

TRogers3@tnstate.edu, jian.peng@ieee.org, mzein@tnstate.edu

***Abstract* - The focus of this paper is the design of a system for Human-Robot Interaction that allows the robot(s) to interact with people in modes that are common to them. The results are a designed architecture for a system to support human-robot interaction. The structure includes a Monitoring Agent for detecting the presence of people, an Interaction Agent to handle choosing robot behaviors that are used for interacting, both socially and for task completion, and a Capability Agent which is responsible for the robot's abilities and actions.**

Index Terms – human-robot interaction, social rules

I. Introduction

This paper reports our work on an implementation of a model for human-robot interaction, the Human Agent System, which was designed to be a flexible, modular representation of interactions between multiple people and a service-oriented robot [1]. The model has been applied to interactions between a stationary humanoid robot and multiple people coming in and out of the robot's environment. In the work reported herein, the Human Agent model is being used to equip mobile robots for human-robot interaction. Several limiting assumptions that had been made are not necessarily valid in the current context. This paper reports the initial findings of applying this model to a mobile robot system, an ActivMedia Pioneer 3 AT robot, see Fig. 1.

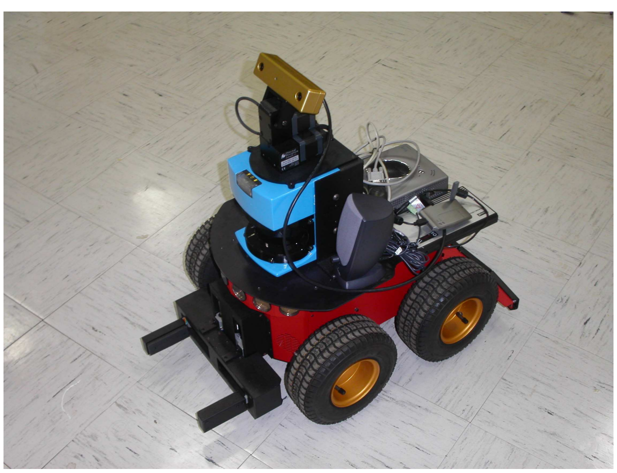

Fig. 1 The Pioneer 3 AT mobile robot

II. Background

Human-robot interaction is a diverse area of study. Much research is conducted with the goal to provide robot systems with the ability to detect and respond to the modes of communication that people use naturally with one another. Social robotics [2] is an area that focuses on the development of robots that operate with people to meet or address some social needs. The field of social robotics is being investigated in two major directions. One focus develops robots with models that give groups of robots social skills for group coordination. A second area of research is investigating specifically how to socially equip robots to respond to needs of people. These needs can include social companionship or entertainment, which try to elicit social responses from people, such as Honda humanoid, toys, Kismet [3], etc. The continuum continues toward the development of systems that draw upon social attitudes to address specific needs of people, such as caregiving in healthcare [4]; autonomous systems such as in response to AAAI Robotics Challenge [5]; and "human-like" personal assistance systems such as ISAC and Cog [6] [7]. This area utilizes studies in interpersonal interaction for application to interactions between people and systems. Studies have shown that people respond to artificial systems with an unconscious similarity to similar interpersonal situations, including a tendency to anthropomorphize or attribute human qualities [8] [9]. Robotics is such a widespread area that the experiences and expectations that people have for their interactions vary widely. It is expected that the use of natural interfaces will allow people to have more realistic expectations of the robot's abilities. The purpose of this paper is to present work applying a model of interaction to the robot.

III. The Human Agent System

The Human Agent System is a model for human-robot interaction based on interpersonal interactions. There are several components of the Human Agent System, which are depicted in Fig. 2. The Human Agent gets input from Human Input Agents and operates on a Human Database. The flexible model has core functionalities that provide a robot with perception, awareness and social appropriateness. These functionalities are described as agents. The term *agent* in this work is used, partially as an artifact of the multi-agent system upon which the Human Agent System was first realized. It is also used to represent the nature of the component modules that function as independent entities with particular functionalities. These agents can then be combined to form higher level agents, and the Human Agent is an example of an agent that is a collective of these more primitive agents.

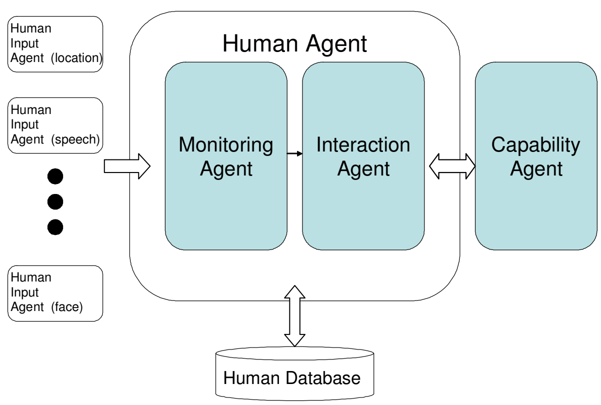

Fig. 2 The Human Agent System

A. Human Agent

The Human Agent is the robot's internal representation of human beings in the environment. It has the responsibility of encapsulating/modeling the human in the environment. It keeps up with the physical, task-related, and cognitive aspects of the human. The Human Agent as described in [1] considers several aspects of interaction. It is composed of several functional modules that perform two key roles. The Monitoring Agent looks for people and people-related events in the environment. The Interaction Agent makes decisions for the robot's participation in interaction, based on its knowledge of social situations and the current state of the interaction environment. The decisions are then forwarded to the portion of the robot that is responsible for high-level task coordination. In this work, the robot's actions and responses are incorporated in a Capability Agent.

B. Monitoring Agent

Perception is represented by a Monitoring Agent, which monitors the environment for features or events that indicate that people are present or active. The monitoring function operates such that the monitor can receive the input from various detectors such as visual (face detection) or environmental (motion).

The approach for human-robot interaction uses several modes for perceiving people. This approach is based on the consideration that interpersonal interaction employs many modes of interaction. It is also founded upon the desire for flexibility when there are limitations on a particular mode of perception. Similar to the work of the Watcher Agent [10], in which cameras monitored the room for a person, the Monitoring Agent monitors the environment for events or features that indicate a person is present.

Human Input Agents (HIAs) are independent processing modules that search for specific features, such as faces or motion, and provide their detection results to the Monitoring Agent. The input from each of these Human Input Agents is connected to one of the three functionalities of the Monitoring Agent. HIAs that perform detection, such as face detection or sound localization are used to *Observe* the environment. The *Identification* functionality is supported by HIAs that perform recognition, which can be based on face, voice or other parameters. The Monitoring Agent design allows for HIAs that perform affect estimation to provide *Human Affect* representation, which will be implemented in future work. This spectrum of features is used to build the representation of the people in the environment.

C. Interaction Agent

The second core feature, which operates on the results of the Monitoring Agent, is responsible for interaction. The Interaction Agent coordinates the robot's role in the interaction. This function processes input and the knowledge of the situation to determine the intention of the people in the environment and determine the appropriate response. The knowledge of the situation can include information such as an express or inferred intention of the people in the environment.

1) Human Intention: The agent creates a description of the intentions of people and uses the current state of the person and attempts to progress to a task. This function process input and knowledge of the situation to determine the intention of the people in the environment and to determine the appropriate response. The knowledge of the situation can include information such as an express or inferred intention of the people in the environment. *Expressed* intentions include the directions or commands a person gives to the robot. *Inferred* intentions are based on what the robot has inferred from the users actions, such as the robot inferring that the particular interaction is over when a person has left the interaction environment.

2) Socialness: The components of the Human Agent System described thus far, provide the robot with the ability to detect features that are relevant to social interaction. However, for the robot to demonstrate sociability, it must act or respond based on this information about the state of the social environment. It is a goal of this work to provide natural interfaces for human-robot interaction.

The appropriateness of an action or response is often determined based on the current social situation. To provide a frame of reference for representing social interactions, the robot is equipped with a model that represents the level of interaction engagement, see Fig. 3. This model of interaction engagement is used to represent an interaction scenario from a state of not being engaged to being fully engaged in an interaction, and can be thought of as states or stages. The pyramid structure is used graphically to represent the desired progression of the interaction from the lower-numbered, foundational levels to higher levels, with the goal of successful task completion.

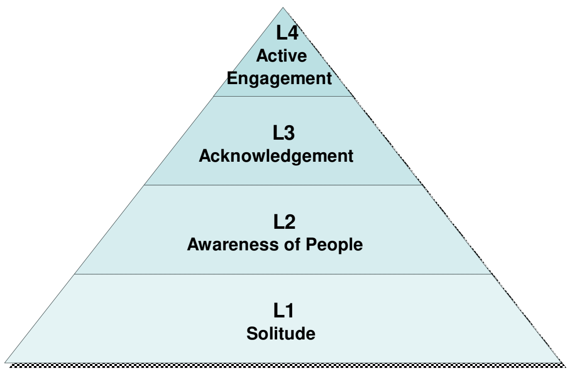

Fig. 3 Levels of Interaction Engagement

A brief description of each of the levels of interaction engagement is presented in Table I. The following levels of interaction engagement are introduced as the frame upon which the Human Agent will navigate to achieve successful task completion.

TABLE I
DESCRIPTIONS OF LEVELS OF INTERACTION ENGAGEMENT

Interaction Level		Interaction Activities
1	Solitude	System is aware of the fact that it is alone and tries to find human interaction.
2	Awareness of people	System may use this state as a trigger to wake up or to initiate interaction.
3	Acknowledgement of person's presence	System utilizes behaviors to set up interaction.
4	Active Engagement	System is actively interacting with a person. Task behaviors occur in this level of engagement.

These levels of interaction engagement are selected by describing the various states that would affect a robot's actions during interaction. The set of levels presented above is influenced by models of confirming responses from interpersonal communication work [11]. These levels are presented as general across interpersonal interactions, and the set of human-robot interactions that are similar to these interpersonal situations. The progression across the interaction levels is based specifically on events that can be viewed as interaction triggers or transitions.

The specific activities at a given interaction level are based on the design goals and capabilities of the robot. For example, a service robot with the goal of providing assistance to people can undertake seeking interaction, based on its physical capabilities, such as wandering around its environment or calling out for people vocally.

Equipping the robot with knowledge about appropriate social responses in various situations allows the robot to choose suitable actions. The robot chooses responses that present it as sociable, which is expected to sustain interaction and encourage the progression of the interaction to full engagement. Table II shows a representative list of general social rules for driving the robot's role in interaction. In the table, these sample rules are grouped by interaction situations.

TABLE II
REPRESENTATIVE SOCIAL RULES GROUPED BY INTERACTION CATEGORY

IN SOLITUDE OR NOT CURRENTLY INTERACTING:
If no person is detected, consider self to be alone and free to do independent or idling behaviors.
• if no people, choose an idle behavior
• if idle > M minutes, change idle behavior
• if idle > N minutes (where N >> M), sleep
GENERAL INTERACTION:
If a new person is detected, behavior is to acknowledge.
• if person is near but does not speak, attempt to engage him
If person leaves, consider interaction ended.
IDENTIFICATION:
If interacting with a new person, identify him.
If know person, relate to other good information about him
If see person for whom there is a relevant or urgent message, deliver it.

D. *Human Database*

A mechanism for maintaining information about interactions, in general and in specific, is handled by a human database. This database logs the history of interactions and can be searched by name or interaction type. It also can hold the features used for personalized interactions, such as identifying features, personal interaction preferences, and pending tasks.

E. *Capability Agent*

The Human Agent System provides the robot with the ability to detect and reason about the state of its human-robot interactions. Its output is the selection of an appropriate or reasonable action or response to the situation. This output is then passed to the Capability Agent to be executed. The Capability Agent is not directly a component of the Human Agent System; however it is critical for robot execution, and therefore, also for the human-robot interaction. The Capability Agent has similarities to the some of the functions of the Self Agent discussed in [12]. The Self Agent, as the robot's internal representation of itself in terms of state and capabilities, was responsible for the robot's high-level decisions about the task performed, as well as the coordination and execution of the task. Its intention resolution capacity allowed the robot to resolve between multiple tasks and interrupting tasks. In the work presented here, it is the role of the Capability Agent to coordinate the actual activities of the robot and to carry out any tasks desired To date the Capability Agent does not handle the higher-level self intention representation that a Self Agent was designed to do. The Capability Agent coordinates and executes a task that the Human Agent System generates. It does not perform conflict resolution between competing tasks.

IV. Implications

The Human Agent System was designed to be flexible and modular, considering the range of human-robot interactions that mainly exist in an environment equivalent to an interpersonal interaction space. The design incorporates the underlying assumption that the model would be designed with its core functionalities (Monitoring and Interaction Agents) and that the specific capabilities of the robot platform would "plug into" this core. For example the Human Agent System was first demonstrated on a humanoid robot with detection abilities that include sound localization, speaker identification, speech recognition, face detection, facial identification, color recognition. It is not the claim of the design that every one of these technologies is critical for the success of the human-robot interaction. It is the specifics of the robot platform that must be considered when moving the model from one robot-environment situation to another. Considerations include mobility and capabilities of the robot, expected environment for interaction, role of robot, etc.

Environmental concerns include the surroundings for the interaction. For example, a noisy environment will place certain constraints if the robot is to listen to the person. The needs for the robot communication will govern design for the interaction. For example, if the robot has to communicate with people who may be hard of hearing, there will be different constraints than if the robot has to prepare to interact with a large multi-lingual audience. The robot's mobility also has an effect on the environment and whether the robot can get into uncertain environments.

V. Implementation

This paper reports the status of progress in the implementation of the Human Agent System on a mobile robot. The robot platform is a Pioneer 3 AT. It has a stereo camera (made by Point Grey Research) mounted on a pan-tilt unit (made by Directed Perception) and a gripper. The gripper can grasp a standing bottle and carry it to a given location.

The software for the model is implemented in Visual Basic using APIs and dlls for modules such as speech recognition, face detection, image processing, etc. In addition to designing the overall structure, two human detection systems, detecting human faces and identifying known speakers by voice, have been implemented. The system has been tested on a library of two people. The work is being integrated into the system for selecting appropriate behaviors based on a history of people.

In this implementation, the Monitoring Agent can detect human faces. A face detector component is employed that is based on the Intel OpenSource Computer Vision library. As the robot searches for people, it looks for faces that are in the range from 1 to 2 meters from the robot. When a face is confidently detected the detector can notify the Human Agent System. The interaction level is updated from Solitude to Awareness of People.

The Interaction Agent operates on the knowledge of people's location, speech, and identity to determine what actions and speech are appropriate. This knowledge also can incorporate history to influence the selection of the next actions. In this context, the robot can process defined speech, such as greetings and task requests. Any decision requiring the robot hardware is then forwarded to the Capability Agent to coordinate its execution.

The task requests, as intentions for the robot to accomplish a task, may also involve activities that must be delayed. For example, one such task is to take an object and store it for another individual. The robot executes the portion of the request that can be done immediately (taking the object to storage). The knowledge of this history allows the robot to recall the previous interactions in which it has participated. The robot can retrieve information about its previous interaction with an individual and extract information that may influence the current interaction.

A. Physical Task

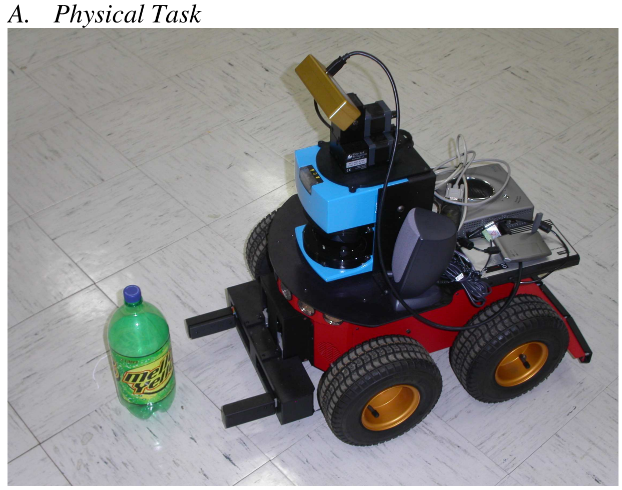

Fig. 4 Robot Preparing to Pick Up a Bottle

The tasks that the robot performs included picking up a bottle that is standing on the floor, taking the object to a storage location, and retrieving objects from the storage, see Fig. 4.

The bottle tasks begin when the robot has been asked to look for a particular (one of three known) type of bottle. When the robot has moved close enough (less than a pre-determined threshold) to the target (bottle), a visual-servoing loop is activated to align the robot's gripper with the bottle and grasp it. During the aligning-and-grasping behavior, the pan-tilt unit points the camera in a fixed pre-determined direction and an active visual servoing is employed. An image-based, endpoint closed-loop control strategy is employed for this aligning-and-grasping behavior [13]. During each iteration of aligning, after a video frame is captured, it goes through color detection, blob finding, and model fitting. Both the target (bottle) and the two fingers of the gripper are detected. This behavior is implemented in two steps, aligning and approaching. The goal of the aligning step is to turn the robot's body so that the bottle aligns in the middle of the two fingers of the gripper. A proportional controller is used, and the rotational velocity ω of the robot is proportional to the difference between the bottle's centroid position (c_{xt}, c_{yt}) and the middle of the gripper fingers on the image (c_{xf}, c_{yf}):

$$\omega = k_{\omega}(c_{xg} - c_{xt}) \tag{1}$$

Once the alignment has finished, the approaching step moves the robot forward by using the following proportional controller:

$$v = k_{v}(c_{yg} - c_{yt}) \tag{2}$$

When the error is less than a pre-defined threshold, the robot stops visual servoing and starts grasping the bottle

B. Deployment

In the implementation and deployment of the Human Agent model, the robot was equipped with the ability to detect people, process and generate speech, and perform tasks for people. The context is interaction between a person and a robot that can perform helper tasks for the person. The robot operates on a limited set of rules based on the current state of the interaction. The robot can greet and identify a person that it has detected. The robot is capable of assisting the person. The particular capabilities are 1) scanning the environment for people, 2) taking an object to a storage location, and 3) retrieving an object from storage.

In the interaction scenario, a robot who is not otherwise engaged will wander about the environment while scanning for people. The robot randomly wanders and then looks around for a human face. When the robot detects a person it can greet and attempt to identify the person. The robot logs interactions and can also check its memory for information of significance that may be related to that person. The robot can then operate on the state of the environment, as well as, any perceived intention of the person.

VI. Results and Conclusions

The results of deploying the robot with the Human Agent System are described below.

A. Solitude

The robot was able to wander about the room when it realized its interaction level to be solitude. In these states the robot either had not begun interacting with a person or had completed its previous interactions. In this state, the robot looked for a person for interaction. The wander behavior allows the robot to operate in response to the social situation. In the solitude state, the robot wanders around its environment, periodically looking for human faces. As its efforts continue to be unsuccessful, it begins to wander in an aggressive mode. In this mode, it searches for people more actively. If searching in the aggressive mode still yields no indication of people in the environment, the robot beings to search more conservatively. This conservative mode provides the potential for the robot detect a person as it returns to an inactive state in its solitude.

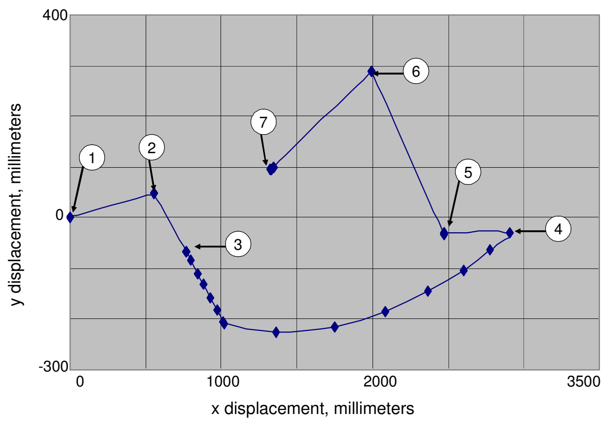

Fig. 5 Robot's Path during Task Execution

Fig 5. shows a trajectory of the robot as it searched for people in its environment. The axes are displacements in millimeters from where it started (0,0). The results included simple interaction of the robot wandering, encountering a person and performing a task. Point 1 is where the robot began its execution. The robot operated in Solitude and began wandering to find people. At point 2, it looked for a person and did not find one. It then continued wandering. At point 3, the robot did become aware of a person when it detected her face. The robot then acknowledged the person. When the person replied, the robot both identified the person based on voice and considered the interaction acknowledged. The robot then offered assistance to the person to attempt to provide service. The person requested that the robot take a specific bottle to the storage area. The robot then went to point 4 and deposited the bottle in storage.

After the requested task was completed, the robot reset its interaction state and again began its Solitude behaviors, wander and search for people. Points 5, 6, and 7 are locations that resulted from the wander commands. At each of these points the robot looked for people by searching for faces. The results were appropriate the environment, i.e. not finding a face where there was not a person (points 5 and 6) and accurately detecting a person at point 7.

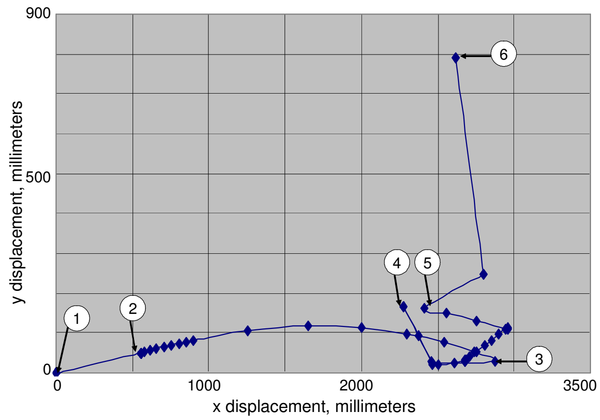

Fig. 6 Robot's Path for Several Interactions

Fig. 6 shows the robot's resulting path during several steps of deployment during a separate trial. In this trial, the system startup occurred at point 1. From this point the robot began its wandering procedure. At point 2 the robot encountered a person. At this point the robot moves from Solitude and progresses to perform the person's task, which is to take the bottle from the floor and place it in storage for a specific person. The robot executes that task, placing the bottle at point 3 and informing of its success. The robot then considered that interaction ended and began to perform a Solitude behavior, wandering. At point 4, the robot encountered, acknowledged, identified and received a task from a second person. In this interaction the robot again progressed from Solitude to Active Engagement. The task requested was to retrieve the particular bottle from storage. The robot then proceeded to the storage area (near point 3), retrieved the bottle and brought it to the person, reaching point 5 and placing the bottle on the floor. After completion of the interaction, the robot began wandering, looking for a person at point 6 and correctly detecting no one there.

One of the interesting considerations that will need to be made as the Human Agent system is developed arose in this trial. In this trial a person requested that the stored bottle be given to a particular individual. Later a person, who was not the person for whom the bottle was stored, requested that the robot retrieve it for him. The current system did not recognize this as a conflict and performed the retrieve bottle task and delivered it to the person. This type of conflict is of the nature that an understanding of the interaction and the nature of task will allow the robot to handle successfully. When placed with a similar situation, a human being would likewise have to consider other factors, potentially including priorities of the individuals and the sensitivity of the object or message to determine if the delivering the object would be appropriate.

C. Conclusions

This work demonstrates the current state of the efforts to place a high-level model for human-robot interaction. It describes the Human Agent System design and flexibility of its implementation domain. With the level of the Human Agent System realized on the mobile robot, we were able to demonstrate the operation of the Solitude level of interaction. The mobility of the robot allows for it to search a greater space for people to serve. The face detector and speaker identifier allowed the robot to transition out of solitude, become aware of people and begin interaction, incorporating personalized history if available.

References

[1] T. Rogers, "The Human Agent: A model for human-robot interaction", Ph. D. Dissertation, Department of Electrical Engineering and Computer Science, Vanderbilt University, 2003.

[2] C. Breazeal, "Designing sociable robots," MIT Press, Cambridge, MA, 2002.

[3] C. Breazeal, "Sociable Machines: Expressive Social Exchange between Humans and Robots", Sc.D. dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2000.

[4] M. Pollack et al., "Pearl: A Mobile Robotic Assistant for the Elderly", *AAAI Workshop on Automation as Eldercare*, Aug., 2002

[5] R. Simmons et al., "GRACE and GEORGE: Autonomous Robots for the AAAI Robot Challenge", *AAAI 2004 Mobile Robot Competition Workshop*

[6] K. Kawamura, T. Rogers, and X. Ao, "Development of a Cognitive Model of Humans in a Multi-Agent Framework for Human-Robot Interaction", *1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AA-MAS)*, Bologna, Italy, July 15-19, 2002.

[7] B. Scassellati, "Theory of Mind for a Humanoid Robot", Ph. D. dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2001.

[8] S. Kiesler and J. Goetz, "Mental Models and Cooperation with Robotic Assistants" in CHI 2002 Extended Abstracts, ACM Press, Minneapolis, MN, April 2002, pp. 576-577.

[9] B. Reeves and C. Nass, The Media Equation, Cambridge University Press, Cambridge, UK, 1996.

[10] H. Takeda, N. Kobayashi, Y. Matusbara, and T. Nishida, "Towards Ubiquitous Human-Robot Interaction", in Working Notes for *IJCAI-97 Workshop on Intelligent Multimodal Systems*, 1997, pp.1-8

[11] K. N. Cissna and E. Sieberg, "Patterns of Interactional Confirmation and Disconfirmation", in Stewart, J. (ed.), Bridges not Walls, McGraw-Hill, New York, 1999

[12] A. Alford, D. M. Wilkes, and K. Kawamura, "System Status Evaluation: Monitoring the State of Agents in a Humanoid System", *IEEE International Conference on Systems, Man and Cybernetics*, Nashville, TN, 2000., pp. 943-948.

[13] J. Peng, A. Peters, X. A. Ao, and A. Srikaew, "Grasping a Waving Object for a Humanoid Robot Using a Biologically-Inspired Active Vision System," presented at *12th IEEE Workshop Robot and Human Interactive Communication*, Silicon Valley, CA, 2003.