

Effects of Nonverbal Communication on Efficiency and Robustness in Human-Robot Teamwork*

Cynthia Breazeal, Cory D. Kidd, Andrea Lockerd Thomaz, Guy Hoffman, Matt Berlin

MIT Media Lab

20 Ames St. E15-449, Cambridge, MA 02139

cynthiab@media.mit.edu

Abstract—Nonverbal communication plays an important role in coordinating teammates’ actions for collaborative activities. In this paper, we explore the impact of non-verbal social cues and behavior on task performance by a human-robot team. We report our results from an experiment where naïve human subjects guide a robot to perform a physical task using speech and gesture. The robot communicates either implicitly through behavior or explicitly through non-verbal social cues. Both self-report via questionnaire and behavioral analysis of video offer evidence to support our hypothesis that implicit non-verbal communication positively impacts human-robot task performance with respect to understandability of the robot, efficiency of task performance, and robustness to errors that arise from miscommunication. Whereas it is already well accepted that social cues enhance the likeability of robots and animated agents, our results offer promising evidence that they can also serve a pragmatic role in improving the effectiveness human-robot teamwork where the robot serves as a cooperative partner.

Index Terms—Human-Robot Interaction, Non-verbal Communication, Teamwork and Collaboration, Humanoid Robots.

I. INTRODUCTION

This work is motivated by our desire to develop effective robot teammates for people. In particular, the issue of how to design communication strategies to support efficient and robust teamwork is very important. In human-human teamwork, sharing information through verbal and non-verbal channels plays an important role in coordinating joint activity. We believe this will be the case for human-robot teams as well.

For instance, Collaborative Discourse Theory specifies the role of dialog in the formulation and execution of shared plans for a common goal [5]. Joint Intention Theory argues that efficient and robust collaboration in dynamic, uncertain, and partially unknowable environments demands an open channel of communication to coordinate teamwork where diverging beliefs and fallible actions among team members are the norm [4]. Much of the existing research has focused on the role of verbal behavior in coordinating joint activity.

Our own work in mixed-initiative human-robot teamwork grounds these theoretical ideas for the case where a human and a humanoid robot work collaboratively to perform a physical task in a shared workspace [3]. Therefore the use of non-verbal behavior in coordinating joint activity plays a very significant, yet relatively understudied role, as compared

to verbal contributions. The focus this work is to better understand the role of non-verbal behavior in coordinating collaborative behavior for physical tasks.

It is important to recognize that non-verbal communication between teammates can be explicit or implicit. We define *explicit* communication as deliberate where the sender has the goal of sharing specific information with the collocutor. For instance, explicit communication transpires when a robot nods its head in response to a human’s query, or points to an object to share information about it with the human. In embodied conversational systems, for instance, explicit non-verbal cues are used by agents to regulate the exchange of speaking turns, convey propositional information, or direct the human’s attention through various gestures and discourse-based facial expressions.

We define *implicit* communication as conveying information that inherent in behavior but which is not deliberately communicated. It is well known that observable behavior can communicate the internal mental states of the individual. Gaze direction can communicate attention and visual awareness, emotive expressions can communicate underlying affective states, and so forth. For example, implicit communication of the robot’s attention transpires when the human reads the robot’s gaze to determine what currently interests the robot.

This paper reports our results from an experiment designed to explore the role and effect of adding implicit non-verbal communication in human-robot teamwork. Naïve human subjects were asked to instruct an autonomous humanoid robot using speech and gesture to perform a simple physical task. The robot does not speak. Instead it communicates non-verbally — either implicitly through behavior or explicitly through gestural social cues. Self-report results via questionnaire offer supportive evidence that implicit non-verbal communication improves transparency of the interaction for the human subject over that of only deliberate non-verbal communication. Behavioral data coded from video of the sessions offers support that the robot’s implicit nonverbal communication improves the efficiency and robustness of the interaction.

II. BENEFITS OF IMPLICIT COMMUNICATION

This paper explores the following three hypotheses regarding how the design of a robot’s implicit non-verbal behavior can benefit the quality of human-robot teamwork.

*This work is funded in part by the *Digital Life* and *Things that Think* consortia of the MIT Media Lab.

Transparency and understandability of the robot’s internal state. We believe that implicit non-verbal communication is important in human-robot teamwork because it conveys *why* the robot behaves as it does. We argue that it makes the robot’s internal state transparent to the human teammate and subsequently more understandable and predictable to her — she intuitively knows how to engage the robot to get the desired result. Given that humans have strong expectations for how particular non-verbal cues reflect specific mental states of another, it is very important that the robot’s implicit non-verbal cues and the internal states to which they map adhere to natural human analogs. This is an important design principle because if they do not, the human is likely to make incorrect inferences about the robot, thereby making the robot’s behavior misleading or confusing to the human.

Efficiency of task performance. We believe that one important outcome of making the robot’s behavior transparent to the human is improved efficiency in task performance. First, by reading these implicit non-verbal cues, the human is better able to fluidly coordinate her actions with those of the robot, potentially saving time and additional steps. These cues can also communicate the robot’s understanding (or lack thereof) to the human without requiring her to request explicit confirmations that take additional time. Third, these cues allow potential sources of misunderstandings to be immediately detected. The human can then quickly adapt her behavior to *preemptively* address these likely sources of errors before they become manifest and require additional steps to correct.

Robustness to errors. Unfortunately, however, errors will occur in human-robot teamwork just as they do in human-human teamwork. We argue that not only is transparency of the robot’s internal state important for improving teamwork efficiency, it also plays an important role in improving teamwork robustness in the face of errors. Implicit non-verbal cues can be used to readily convey to the human *why* an error occurred for the robot, often due to miscommunication. This allows her to quickly address the *correct* source of the misunderstanding to get the interaction quickly back on track. Otherwise misunderstandings shall persist until correctly identified and could continue to adversely impact the interaction.

III. RELATED WORK

Whereas past research has shown that non-verbal social cues improve the likeability of robots and interactive characters, demonstrating their ability to effect improved task performance has been elusive. In past work, the embodied agent (often virtual) usually acts as an assistant or advisor to a human in solving an information task. In our scenario, the human leads the interaction but she is dependent on the robot to do the actual work. This level of interdependence between human and robot may make communication between them sufficiently important to be able to see the effects of implicit non-verbal behavior on task performance, and particularly its role in coordinating joint action.

In addition, this study investigates the impact of the robot’s physical and non-verbal behavior on the human’s mental model for the robot, in contrast to prior works that have explored how this mental model is influenced by what the robot looks like (i.e., its morphology) or its use of language (e.g., [9]).

We are not aware of Human-Robot Interaction (HRI) studies that have systematically explored the issue of teamwork robustness in the face of errors where the robot is completely autonomous (e.g., rather than teleoperated as in USAR work). For instance, many HRI experiments adhere to a Wizard of Oz methodology to bypass the physical and cognitive limits of what robots can do today (e.g., [6]). This is done for good reasons, but it misses the opportunity to investigate how to design autonomous robots that successfully mitigate errors that inevitably do arise in human-robot teamwork to do common performance limitations.

In contrast, our robot runs completely autonomously, and therefore is subject to making typical errors due to limitations in existing speech recognition and visual perception technologies. For instance, the human subjects in our study speak with different accents and at different speeds. They wear clothes or stand at interpersonal distances from the robot that can adversely affect the performance our gesture recognition system. This gives us the opportunity to systematically investigate how to design communication cues to support robust human-robot teamwork in the face of these typical sources of miscommunication.

Finally, in HRI studies where the robot operates completely autonomously, the interaction is typically robot-lead (e.g., [14]). This allows researchers to design tasks, such as information sharing tasks or hosting activities, where the human’s participation can be restricted to stay within the robot’s performance limitations (such as only being able to give “yes” or “no” responses to a robot’s queries). In contrast, this work explores a human-lead task.

Consequently, our human subjects have significant flexibility and show substantial variability in how they interact with the robot to perform the task. For instance, as mentioned above, people speak differently, wear different clothes, and choose to stand different distances from the robot. The style of their gestures also varies widely, and they each accomplish the task using a different series of utterances. This places higher demands on the robot to respond dynamically to the human’s initiatives. However, our task is structured sufficiently (in contrast to more freeform interaction studies as in [8]) to be able to compare task performance across subjects for different conditions. This allows us to investigate how human behavior varies along important dimensions that impact teamwork performance.

IV. EXPERIMENTAL PLATFORM

Our research platform is Leonardo (“Leo,” See Fig. 1), a 65 degree of freedom expressive humanoid robot designed for social interaction and communication to support teamwork [7] and social learning [11]. The robot has both speech-based

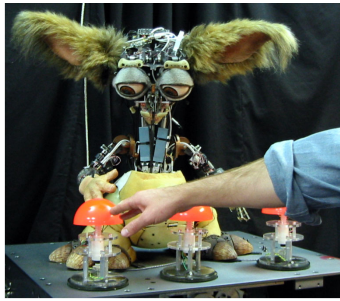


Fig. 1. Leo and his workspace with three buttons and a human partner.

and visual inputs. Several camera systems are used to parse people and objects from the visual scene [2].

In the task scenario for this experiment, the human stands across the workspace facing the robot. A room-facing stereo-vision system segments the person from the background, and a Viola-Jones face detector is used to locate her face. A downward facing stereo-vision system locates three colored buttons (red, green and blue) in the workspace. It is also used to recognize the human’s pointing gestures. A spatial reasoning system is used to determine to which button the human is pointing. The speech understanding system, implemented using Sphinx-4 [10], uses a limited grammar to parse incoming phrases. These include simple greetings, labeling the buttons in the workspace, requesting or commanding the robot to press or point to the labeled buttons, and acknowledging that the task is complete.

These speech-related and visual features are sent to the cognitive system (an extension of the C5M architecture [1] that models cognitive processes such as visual attention, working memory, and behavior arbitration) where they are bundled into coherent beliefs about objects in the world and communicated human intentions, which are then used to decide what action to perform next. These actions include responding with explicit non-verbal social cues (e.g., gestures and communicative expressions as shown in Table I), as well as task-oriented behaviors with implicit communicative value — such as directing attention to the relevant stimuli, or pressing the buttons ON or OFF. The cognitive system also supports simple associative learning, such as attaching a label to an object belief, allowing the human to teach the robot the names of objects in its workspace.

V. EXPERIMENT

Our experiment is designed to test the effects of Leo’s nonverbal expressions in cooperative interactions with naïve human subjects. Each subject was asked to guide the robot through a simple button task where the subjects first taught the robot the names of the buttons, and then had the robot turn them all on. Although simple, this scenario is sufficiently rich in that it provides opportunities for errors to occur. Specifically, there are two potential sources of errors in communication:

- The gesture recognition system occasionally fails to recognize a pointing gesture. Or,
- The speech understanding system occasionally misclassifies an utterance.

Furthermore, errors that occur in the first part of the task (the labeling phase) will cause problems in the second part of the task (the button activation phase) if allowed to go undetected or uncorrected. In addition, the robot also suffers occasional glitches in its behavior if a software process crashes unexpectedly. If this malfunctioning prevented the human subject from completing the task, their data was discarded.

A. Manipulations

Two cases are considered in this experiment. In the **IMP+EXP** case, the robot pro-actively communicates internal states implicitly through non-verbal behavior as well as explicitly using expressive social cues. In the **EXPLICIT** case, the robot only explicitly communicates these internal states when prompted by the human. This manipulation allows us to investigate the added benefit of implicit non-verbal communication over above explicit non-verbal communication which as been more widely investigated (e.g., in embodied conversational agents).

For instance, in the **IMP+EXP** case (Table I), nonverbal cues communicate the robot’s attentional state to the buttons and to the human through changes in gaze direction in response to pointing gestures, tracking the human’s head, or looking to a particular button before pressing or pointing to it. In addition, the robot conveys liveliness and general awareness through eye blinks, shifts in gaze, and shifts in body posture between specific actions. Its shrugging gestures and questioning facial expression conveys confusion (i.e., when a label command does not co-occur with a pointing gesture, when a request is made for an unknown object, or when speech is unrecognized). Finally, the robot replies with head nods or shakes in response to direct yes/no questions, followed by demonstration if appropriate.

The **EXPLICIT** case, in contrast, removes the implicit cues that reveal the robot’s internal state. Eye gaze does not convey the robot’s ongoing attentional focus in response to the human. Instead, the robot looks straight ahead, but will still look at a specific button preceding a press or point action. There are no behaviors that convey liveliness. The robot does not pro-actively express confusion, and only responds with head nods and shakes to direct questions.

B. Procedure

A total of 21 subjects were drawn from the local campus population via e-mail announcements. Subjects were nearly evenly mixed in gender (10 males, 11 females) and ranged in age from approximately 20 to 40 years. None of the participants had interacted with the robot before.

Subjects were first introduced to Leo by the experimenter. The experimenter pointed out some of the capabilities of the robot (such as pointing and pressing the buttons) and indicated

TABLE I
 IMPLICIT CASE: WITH BEHAVIORAL AND NONVERBAL CUES

| Context | Leo's Expression | Intention |
|---------------------------------------|--------------------|-----------------------------------|
| Human points to object | Looks at object | Shows object of attention |
| Human present in workspace | Gaze follows human | Shows social engagement |
| Human asks yes/no question | Nod/Shake | Communicates knowledge or ability |
| Human greets robot | Nod | Issues greeting |
| End of task | Nod | Communicates task is complete |
| Label command has no pointing gesture | Confusion gesture | Communicates problem to human |
| Request is made for an unknown object | Confusion gesture | Communicates problem to human |
| Speech did not parse | Confusion gesture | Communicates problem to human |
| Between requested actions | Idle body motion | Creates aliveness |
| Intermittent | Eye blinks | Creates aliveness |
| Intermittent | Shifts in gaze | Conveys awareness |

a list of example phrases that the robot understands. These phrases were listed on a series of signs mounted behind the robot. The subject was instructed to complete the following button task with the robot.

- Teach Leo the names and locations of the buttons.
- Check to see that the robot knows them.
- Have Leo turn on all of the buttons. And,
- Tell Leo that the "all the buttons on task" is done.

After the task, a questionnaire was administered to the subject. After completion, the subject could choose whether or not to interact with the robot again. If the subject decided to continue, they were asked to try to teach the robot a new task, and example phrases were given for how this could be done.

C. Hypotheses and Measures

The questionnaire covered several topics such as the readability and transparency of Leo's actions and expressions; the subject's mental model of the interaction; and the perceived effectiveness of the interaction. On these topics we have three hypotheses (H1-H3):

- H1: Subjects are better able to understand the robot's current state and abilities in the IMP+EXP case.**
- H2: Subjects have a better mental model of the robot in the IMP+EXP case.**
- H3: The interaction is viewed as more effective from the subject's point of view in the IMP+EXP case.**

In addition to the questionnaire data, each session was video recorded. We have three more hypotheses (H4-H6) related to the behavioral observations from this data. From the video we had the following measures coded: the total number of errors during the interaction; the time from when an error occurred to being detected by the human; the length of the interaction as measured by time and by the number of utterances required to complete the task. These measures test the following hypotheses:

- H4: The total length of the interaction will be shorter in the IMP+EXP case.**
- H5: Errors will be more quickly detected in the IMP+EXP case.**
- H6: The occurrence of errors will be better mitigated in the IMP+EXP case.**

VI. RESULTS

A. Questionnaire Results

In the questionnaire, two of our hypotheses were confirmed. There was a significant difference between the two manipulations on answers to questions about subject's ability to understand the robot's current state and abilities. Thus Hypothesis 1 is confirmed and people perceived that the robot was more understandable in the IMP+EXP case: $t(11) = -1.88$, $p < 0.05$.

There was also a significant difference from the questions concerning the subject's mental model of the robot (e.g. "Was it clear when the robot was confused?", "Was it clear when it understood what I had referred to?", etc.). This confirms Hypothesis 2, that the subjects perceived they had a better mental model of the robot in the IMP+EXP case: $t(11) = -1.77$, $p = 0.05$.

The implicit non-verbal communication had no effect on whether or not subjects reported the interaction to have been effective (Hypothesis 3). We do, however, have indications that the behavioral data supports this claim.

B. Behavioral Results

Our video analysis offers very encouraging support for Hypotheses 4 through 6. Of the 21 subjects, video of 3 subjects was discarded. In two of these discarded cases, the robot was malfunctioning to the point where the subjects could not complete the task. In the remaining case, the subject lost track of the task and spent an unusually long time playing with the robot before she resumed the task. Therefore, the video was analyzed for a total of 18 subjects, 9 for the IMP+EXP case and 9 for the EXPLICIT case. Table II summarizes the timing and error results of the video coding.

TABLE II
TIME TO COMPLETE THE TASK FOR EACH CASE AS A FUNCTION OF THE
NUMBER OF ERRORS (e).

| Condition | Category | Errors | Avg Task Time (sec) |
|-----------|----------------------|--------|---------------------|
| IMP+EXP | all samples | avg=3 | 101 |
| | A: $e \leq 1$ | max=1 | 64 |
| | B: $2 \leq e \leq 4$ | max=3 | 119 |
| | C: $e > 4$ | max=6 | 118 |
| EXPLICIT | all samples | avg=6 | 175 |
| | A: $e \leq 1$ | max=1 | 82 |
| | B: $2 \leq e \leq 4$ | max=4 | 184 |
| | C: $e > 4$ | max=11 | 401 |

On average, the total time to complete the button task was shorter for the IMP+EXP case, offering support for Hypothesis 4. The average time for the subjects to complete the task in the IMP+EXP case is 101 seconds, versus 175 seconds in the EXPLICIT case. By breaking each case into three categories, based on the number of errors that transpired during the interaction (category A: $e \leq 1$, category B: $2 \leq e \leq 4$, and category C: $e > 4$), we see that the IMP+EXP case took less time to complete in each category, with a more dramatic difference in time for each category as the number of errors increased — category A: IMP+EXP=64 vs. EXPLICIT=82; category B: IMP+EXP=119 vs. EXPLICIT=184; category C: IMP+EXP=118 vs. EXPLICIT=401. Analyzing only those trials where at least one error occurred, the average task time for the IMP+EXP case was 107 seconds with a standard deviation of 53.8. In contrast, the average task time for the EXPLICIT case where at least one error occurred was 246 seconds (over twice as long), with a standard deviation of 159.6 (over twice as large).

From video analysis, errors were more quickly detected in the IMP+EXP case, supporting Hypothesis 5. As stated earlier, there were two common sources of error in communication. First, the gesture recognition system occasionally fails to recognize a point gesture. This could be due to several factors, such as the clothes the subject was wearing (long sleeves interfered with skin-tone segmentation), standing far from the robot so that their hand was far from the buttons when pointing to them, standing very close to the robot so that the pointing gesture was cramped, or making the pointing gesture too quickly for the system to reliably register it. This is readily apparent to the subjects in the IMP+EXP case because the robot fails to look at the intended button. Because the robot’s gaze does not reflect its attentional state in the EXPLICIT condition, the subject do not find out that the robot failed to acquire the correct label for a particular button until explicitly asked to do something with that button (e.g., point to it or press it). It is important to note that all subjects naturally wanted to rely on the robot’s gaze behavior as a cue to the robot’s attentional state. Subjects in the EXPLICIT case often looked a bit confused when the robot did not visually track their pointing gesture, and often made a concerted effort to look into the robot’s eyes to see if it was visually responsive.

The second common source of error arose when the speech understanding system misclassifies an utterance. This error was immediately detected in the IMP+EXP case because the robot pro-actively displays an expression of confusion when a speech-related error occurs. In the EXPLICIT case, the robot does not express it’s internal state of “confusion,” and therefore the subjects could not tell whether the robot understood them and was taking an unusually long time to respond, it simply missed its turn, or it failed to understand their utterance. As a result, the EXPLICIT case had varying numbers of awkward pauses in the interaction depending on how well the speech recognition system could handle the subject’s speaking style.

Finally, the occurrence of errors appears to be better mitigated in the IMP+EXP case. On average, it took less time to complete the task and fewer errors occurred in the IMP+EXP case. For the EXPLICIT case, the standard deviation over the number errors (excluding the error-free trials) is over twice as large as that of the IMP+EXP case, indicating less ability to mitigate them in the EXPLICIT case. As can be seen in category C, almost twice as many errors occurred in the EXPLICIT case than in the IMP+EXP case. Video analysis of behavior suggests that the primary reason for this difference is that the subjects had a much better mental model of the robot in the IMP+EXP case due to the non-verbal cues used to communicate the robot’s attentional state and when it was “confused.” As a result, the subjects could quickly see when a *potential* error was *about to occur* and they quickly acted to address it.

For instance, in the IMP+EXP case, if the subject wanted to label the blue button and saw the robot fix its gaze on the red button not shift it over to the blue one, she would quickly point to and label the red button instead. This made it much more likely for the robot to assign the correct label to each button if the perception system was not immediately responsive. In addition, in the IMP+EXP case, the subjects tightly coordinated their pointing gesture with the robot’s visual gaze behavior. They would tend to hold their gesture until the robot looked at the desired button, and then would drop the gesture when the robot initiated eye contact with them, signaling that it read the gesture, acquired the label, and was relinquishing its turn. It is interesting to note that even in IMP+EXP category C where a number of errors were made, the time to complete the button task was very similar to IMP+EXP category B. This offers support that errors that occurred in the IMP+EXP case were quickly detected and repaired so that the overall task time was not dramatically adversely affected.

VII. DISCUSSION

This experiment investigates a cooperative social interaction between a human and a robot. Our results illustrate the importance and benefit of having a robot’s implicit and explicit non-verbal cues adhere to fundamental design principles of the psychology of design [12], [13]. Specifically, we observed that the design principles of *feedback*, *affordances*,

causality, and *natural mappings* play a critical role in helping naive human subjects maintain an accurate mental model of the robot during a cooperative interaction. This paper shows the effectiveness of these basic design principles when adapted to the social interaction domain. People certainly relied on their mental model to interact with the robot, and our data indicates that they were better able to cooperate with Leonardo when they could form a more accurate mental model of the robot.

For instance, the robot pro-actively provides feedback in the IMP+EXP case when it shrugs in response to failing to understand the person’s utterance. This immediately cues the human that there is a problem that needs to be corrected. The robot’s eyes afford a “window” to its visual awareness, and having the robot immediately look to what the human points to signals to her that her gesture causes the robot to share attention — confirming that her intent was correctly communicated to and understood by the robot. Leonardo’s explicit non-verbal cues adhere to natural mappings of human non-verbal communication, making them intuitive for the human to understand. For instance, having Leonardo re-establish eye contact with the human when it finishes its turn communicates that it is ready to proceed to the next step in the task.

We also found that the social cues of a robot should carefully adhere to these design principles otherwise the robot’s behavior becomes confusing or even misleading. For instance, in one trial the robot was accidentally giving false cues. It nodded after a labeling activity, which was a spurious action, but led the human to believe that it was acknowledging the label. As a result, it took the human a longer time than usual to figure out that the robot had actually not acquired the label for that button.

When these cues allowed the human to maintain an accurate mental model of the robot, the quality of teamwork was improved. This transparency allowed the human to better coordinate her activities with those of the robot, either to foster efficiency or to mitigate errors. As a result, the IMP+EXP case demonstrated better task efficiency and robustness to errors. For instance, in viewing the experimental data, the subjects tend start off making similar mistakes in either condition. In the IMP+EXP condition, there is immediate feedback from Leonardo, which allows the user to quickly modify their behavior, much as people rapidly adapt to one another in conversation. In the EXPLICIT case, however, subjects only receive feedback from the robot when attempting to have him perform an action. If there was an error earlier in the interaction that becomes manifest at this point, it is cognitively more difficult to determine what the error is. In this case, the expressive feedback in the IMP+EXP condition supports rapid error correction in training the robot.

VIII. CONCLUSION

The results from this study informs research in human-robot teamwork [7]. In particular, this study shows how people read and interpret non-verbal cues from a robot in

order to coordinate their behavior in a way that improves teamwork efficiency and robustness. We found that people infer task-relevant “mental” states of Leonardo not only from explicit social cues that are specifically intended to communicate information to the human (e.g., nods of the head, deictic gestures, etc), but also from implicit behavior (e.g., how the robot moves its eyes: where it looks and when it makes eye contact with the human). Furthermore, they do so in a consistent manner with respect to how they read and interpret the same non-verbal cues from other humans.

Given this, it is important to appreciate that people have very strong expectations for how implicit and explicit non-verbal cues map to “mental” states and their subsequent influence on behavior and understanding. These social expectations need to be supported when designing human-compatible teamwork skills for robots. This is important for anthropomorphic robots such as humanoids or mobile robots equipped with faces and eyes. However, we believe that in any social interaction where a robot cooperates with the human as a partner, people will want these cues from their robot teammate. If the robot provides them well, the human will readily use them to improve the quality of teamwork.

In the future, robot teammates should return the favor. In related work, we are also exploring how a robot could read these same sorts of cues from a human, to better coordinate its behavior with that of the human to improve teamwork. This shall be particularly important for performing cooperative tasks with humans in dynamic and uncertain environments, where communication play a very important role in coordinating cooperative activity.

ACKNOWLEDGMENTS

This work is funded in part by the *Digital Life* and *Things that Think* consortia of the MIT Media Lab. The work presented in this paper is a result of the ongoing efforts of the graduate and undergraduate students of the MIT Media Lab Robotic Life Group and our collaborators. Particular thanks go out to Jesse Gray for his involvement in this project. Stan Winston Studio provided the physical Leonardo robot. Geoff Beatty and Ryan Kavanaugh provided the virtual model and animations. Leonardo’s architecture is built on top of the C5M code base of the Synthetic Character Group at the MIT Media Lab, directed by Bruce Blumberg.

REFERENCES

- [1] B. Blumberg, R. Burke, D. Isla, M. Downie, and Y. Ivanov. CreatureSmarts: The art and architecture of a virtual brain. In *Proceedings of the Game Developers Conference*, pages 147–166, 2001.
- [2] C. Breazeal, A. Brooks, D. Chilongo, J. Gray, G. Hoffman, C. Kidd, H. Lee, J. Lieberman, and A. Lockerd. Working collaboratively with humanoid robots. In *Proceedings of IEEE-RAS/RSJ International Conference on Humanoid Robots (Humanoids 2004)*, Santa Monica, CA, 2004.
- [3] C. Breazeal, G. Hoffman, and A. Lockerd. Teaching and working with robots as a collaboration. In *Proceedings of Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS04)*, pages 1030–1037, New York, NY, 2004.
- [4] P. Cohen and H. Levesque. Teamwork. *Nous*, 25:487–512, 1991.
- [5] B. J. Grosz. Collaborative systems. *AI Magazine*, 17(2):67–85, 1996.

- [6] P. Hinds, T. Roberts, and H. Jones. Whose job is it anyway? a study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, 19:151–181, 2004.
- [7] G. Hoffman and C. Breazeal. Collaboration in human-robot teams. In *Proc. of the AIAA 1st Intelligent Systems Technical Conference*, Chicago, IL, USA, September 2004. AIAA.
- [8] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro. Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, 19:61–84, 2004.
- [9] S. Kiesler and J. Goetz. Mental models of robotic assistants. In *Proceedings of Conference on Human Factors in Computing Systems (CHI2002)*, pages 576–577, Minneapolis, MN, 2002.
- [10] P. Lamere, P. Kwok, W. Walker, E. Gouvea, R. Singh, B. Raj, and P. Wolf. Design of the cmu sphinx-4 decoder. In *8th European Conf. on Speech Communication and Technology (EUROSPEECH 2003)*, 2003.
- [11] A. Lockerd and C. Breazeal. Tutelage and socially guided robot learning. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2004.
- [12] D. Norman. *The Psychology of Everyday Things*. Basic Books, New York, 1988.
- [13] D. Norman. How might humans interact with robots. Keynote address to the DARPA/NSF Workshop on Human-Robot Interaction, San Luis Obispo, CA, September 2001.
- [14] C. Sidner and C. Lee. Engagement rules for human-robot collaborative interaction. *IEEE International Conference on Systems, Man and Cybernetics*, 4:3957–3962, 2003.