2010

The Effect of Gesture on Collaboration with an Autonomous Robot

by

Kevin O'Brien

Joel Sutherland

This report represents the work of two WPI undergraduate students submitted to the faculty as evidence of completion of a degree requirement. WPI routinely publishes these reports on its web site without editorial or peer review.

The Effect of Gesture on Collaboration with an Autonomous Robot

A Major Qualifying Project Report

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

by

____

Kevin O'Brien (Computer Science)

____

Joel Sutherland (Robotics Engineering)

Approved:

____

Prof. Charles Rich, Major Advisor

Abstract

The goal of this project was to study the effect of robot gestures on the quality of collaboration between an autonomous robot and a human. We conducted a spoken-language human-robot interaction study with 26 participants which showed that adding robot head turns (to make eye contact with the human) and robot arm gestures (to point toward task-related objects) improved the human's perception of the robot's interaction skills and role as a good collaborator. These results have important implications for the design of human-robot interactions in general.

Acknowledgements

We would like to thank Professor Rich as well as Professor Candy Sidner and Professor Joe Beck for their expertise and support. This work is supported, in part, by the National Science Foundation under award IIS-0811942.

Table of Contents

Table of Figures

Chapter 1: Introduction

Robots have the potential to automate many tasks in the future, from selling books to informing a passerby where the nearest hovercraft station is. For robots to do this effectively, they will need to communicate with humans and be able to receive information from them, interpret it correctly and return useful feedback. They must also be able to collaborate, which involves much more than just processing information and speaking. One very important aspect of human collaboration is the use of gestures for communication and engagement. Many studies have been conducted regarding speech and audio in human-robot interaction. However, there needs to be more research on the role of gestures in human-robot interaction in order to improve collaboration and make conversations feel more natural to humans.

To make conversations with robots feel more natural, we apply knowledge and concepts from human-human interaction to human-robot interaction. It is generally known that gestures contribute a lot to human communication. For example, gaze can be used to establish eye contact with the person you are talking to, which indicates that you want them to listen to you. This is a common method for keeping an audience engaged during public speeches. Looking or pointing at objects can help convey information about where the attention of the addressee should be in order to better understand the conversation. In our study, we studied the effects of two types of gestures on human-robot interaction: head turns (a type of gaze) and pointing toward task-related objects.

1.1 Background

Gaze and pointing are two of the most studied types of gesture in human interaction. Argyle and Cook [1976] documented the function of gaze as an overall social signal and noted that failure to attend to another person via gaze is evidence of lack of interest and attention. Other researchers have offered evidence of the role of gaze in coordinating talk between speakers and hearers, in particular, how gestures direct gaze to the face and why gestures might direct gaze away from the face [Kendon 1967, Duncan 1972, Goodwin 1986]. Nakano et al. [2003] reported on the use of the hearer's gaze and the lack of negative feedback to determine whether the hearer has grounded [Clark 1996] the speaker's turn.

Human gaze also serves to facilitate the learning process and enhance task performance. College students had higher performance on a learning task when the instructor gazed at them than when the instructor did not [Fry & Smith 1975]. Furthermore, when students are able to return the gaze to the instructor, they participate more in the instruction than when they are not able to gaze at the instructor [Caproni et al. 1977]. Situations that foster mutual gaze are especially preferred during cooperation tasks when multiple individuals are working towards a goal [Jellison & Ickes 1974]. Most research on human pointing behavior, such as [Jarvella & Klein 1982] is concerned with how pointing contributes to reference resolution and intent recognition.

1.2 Related Work

The previous work mostly closely related to this project is by Sidner et al. [2005], who performed a study that looked into the role of head nod gestures in what they called "hosting activities". A hosting activity is where the host (in this case, a robot) provides another agent (a human) with information, entertainment or other services typically involving their shared environment. In Sidner's study, the robot asks the human to help it complete a series of collaborative tasks. In one condition, the robot uses head nods; in the other it does not.

The first question Sidner studied was whether or not people respond to robots in the same way they respond to other humans. To develop this further, she videotaped interactions between two people and marked occasions where either participant used a head nod. She then analyzed the meaning behind these head nods to understand the role they played in the communication scenario. The data from this observation resulted in the Principle of Conversational Tracking: a participant in a collaborative conversation tracks the other participant's face during the conversation in balance with the requirement to look away in order to: (1) participate in actions relevant to the collaboration, or (2) multitask with activities unrelated to the current collaboration. This principle was used in their experiment by allowing the robot to move its head towards actions relevant to the conversation, but not towards ones unrelated.

Sidner et al.'s experiment involved a robot that instructed participants to pour water into a glass. In one condition, the robot would turn its head towards the participant when talking to him and use head nods. In the other condition, the robot would only move its beak when talking and would not perform any head nods, looking straight ahead throughout the experiment. Participants with the first condition felt as though they were more engaged, directing their attention to the robot more often.

Chapter 2: Method

2.1 Study Concept

The purpose of this project was to study the impact of gestures on human-robot collaboration. We devised an experiment in order to evaluate how collaboration could be improved when gestures were present. We needed an environment and an interaction which would provide the opportunity for study participants to become involved and engaged so that the addition of gestures would have the greatest effect.

Figure 1: Humanoid Robot RSMedia

We began by determining that an autonomous humanoid robot (shown in Figure 1) would provide the best partner for the study subjects to interact with. Such a robot would be the closest approximation to another human, so we would have the best chance of study participants allowing themselves to interact with the system in a natural, human way. It was also important that the robot be capable of speaking back to the participant, as a speech interface is the most natural way for a human to communicate.

Story-Driven Scenario

To further encourage the study participants to become engaged, we decided to develop a story-driven scenario with specific objectives for the study participant and the robot to accomplish as a team. The story gave the robot a name ("Larry") and a human voice to further the impact. The idea behind this design is that if the participant is involved and paying attention to the story, they are less distracted by the study itself. We decided to implement an “obstacle course” design which involves several opportunities for collaboration and communication between human and robot. Larry walks up to an obstacle, describes what it and then proposes a solution to getting through it. The solution will involve the human helping by retrieving something, moving something or otherwise interacting with Larry and the environment. Larry begins the interaction by introducing himself and informing his human collaborator of the major goal he wants help to achieve. Along the way Larry comes across a number of obstacles. Overcoming them are the subgoals which contribute the ultimate goal, completing the scenario. See Appendix B for the complete scenario script.

2.2 Study Design

Conditions

We implemented two conditions for our study: a control condition, and a “gesture” condition. In the gesture condition, the following gestures were implemented:

1. The robot turns its head to look at the participant before speaking.
2. The robot gestures in the direction of objects the participant needs to use.

The control condition did not have these gestures.

Hypotheses

In preparation for our study, we formulated the following five hypotheses. We predicted that in the gesture condition:

1. Participants will think the robot has better interaction skills
2. Participants will think the robot is a better collaborator.
3. Participants will enjoy the interaction more.
4. Completing the goal of the interaction will take less time
5. Participants will perceive that completing the goal of the interaction took less time

Objective and Subjective Measures

Completion Time

The only objective measure in our study was the completion time. In order to evaluate Hypothesis 4, that the interaction would take less time in the gesture condition, we videotaped the interaction and later extracted the time it took to complete the interaction.

A related subjective measure we decided to include was the *perceived* completion time. We anticipated that study subjects would enjoy the interaction more, and thus perceive that the interaction took less time than it actually did (Hypothesis 5).

Questionnaire Design

To evaluate Hypotheses 1 through 3, we created a questionnaire with 25 questions in three categories: (1) Interaction Skills (2) Collaboration, and (3) User Experience. These three categories directly correspond to the aforementioned hypotheses. To get more reliable information, multiple questions within each category all addressed the same issue, but were worded differently. For example "I felt that Larry understood me" and "I could tell when Larry did not understand me" are two questions that ask for the same kind of information about interaction skills, but are worded differently. The question categories are explained in more detail below. The full questionnaire is shown in Appendix F.

1. <u>Interaction Skills</u>

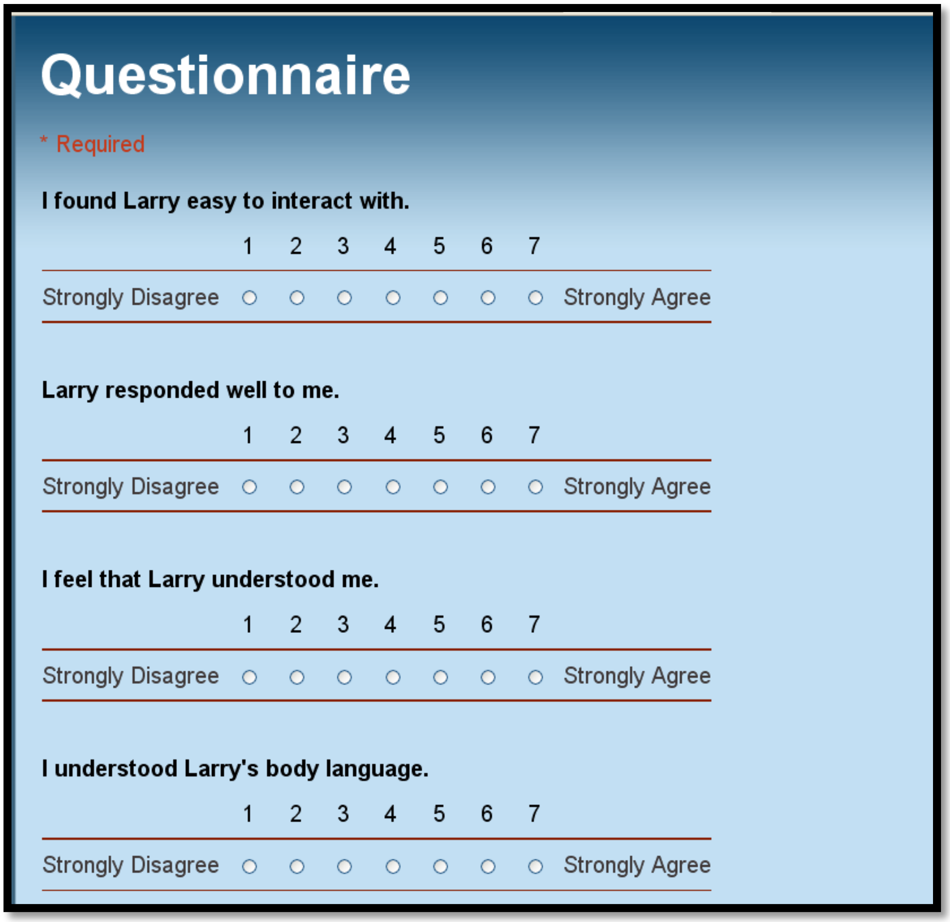
   This category contains 14 questions which queried participants regarding how they felt about the effectiveness of the interaction. This includes finding out how well they understood Larry, how well they felt Larry understood them, and how confused they were during the interaction. "Larry responded well to me" is an example question.

2. <u>Collaboration</u>

   This category contains 6 questions related to parts of the interaction involving joint efforts between Larry and the participant. Questions addressed how the participants felt about how well they worked with Larry, whether or not Larry was a good partner, dependence on Larry and engagement. "Larry and I worked well together" is an example of one of the questions addressing collaboration.

3. <u>User Experience</u>

   This category contains 5 questions dealing with how well the participant enjoyed the interaction, how involved in the scenario they were, and how comfortable they felt during the interaction. A sample question is "I felt comfortable during the task".


Questionnaire

* Required

I found Larry easy to interact with.

	1	2	3	4	5	6	7	
Strongly Disagree	○	○	○	○	○	○	○	Strongly Agree

Larry responded well to me.

	1	2	3	4	5	6	7	
Strongly Disagree	○	○	○	○	○	○	○	Strongly Agree

I feel that Larry understood me.

	1	2	3	4	5	6	7	
Strongly Disagree	○	○	○	○	○	○	○	Strongly Agree

I understood Larry's body language.

	1	2	3	4	5	6	7	
Strongly Disagree	○	○	○	○	○	○	○	Strongly Agree

Figure 2: Questionnaire Screenshot

We used an online questionnaire developed in Google Documents (docs.google.com), to make the data retrieval and analysis process more efficient. Figure 2 shows exactly what the participants saw when they were directed to the computer at the end of the scenario. Even the web browser window was hidden from view in order to keep the participant focused and not confused. We used the 7-point Likert scale, which is most commonly used for this kind of subjective questioning. The data was collected automatically in a spreadsheet format that was easy to export to Excel to make charts, graphs and apply statistical tests.

Preparation and Setup

When each study participant arrived, he or she was first asked to sit at a table outside the experiment room. The participant was given the informed consent form (see Appendix C) to read and sign if he or she decided to participate in the study. Each participant was also given the option not to be videotaped. After the form was signed, the participant entered the room, was seated and given further instructions (see below). There were a number of steps that needed to be implemented by the experimenters in order to set up each participant's interaction correctly. In more detail, here are the preparation and setup procedures we used:

Procedure

Before Participant Enters

- Ensure participant has signed the informed consent form
- Change robot batteries if needed (every third or fourth run)
- Initialize Larry
    - Turn on
    - Initialize to appropriate condition
    - Turn off eyes and ears ('X' and 'C')
    - Sleep Mode (L+Stop)

When Participant Enters

- Seat participant in chair
- Give participant command sheet (see Appendix E)
- Turn on microphone amplifier
- Set up headset
    - Put headset in proper position (and instruct subject)
    - Place microphone correctly (and instruct subject)
    - Conduct IR tests
        - Disable IR Transmit
        - Have subject speak each command from the cheat sheet
        - Verify that each was recognized
        - Show the subject the IR transmitter and tell him not to block it
- Tell the participants
    - To never pick up or grab Larry for any reason
    - When they are finished they will need to fill out the questionnaire, and that Larry will remind them to do so
    - Wait a few seconds before talking to Larry after he finished
    - Say "Affirmative" when ready
    - Turn off cell phone
    - Tell them they will have to stand up and should be careful with the microphone cable
- Re-enable IR Transmit
- Start video recording
- Leave the room

When Subject Exits Room

- Direct them to the refreshments
- Stop video recording
- Turn amplifier off
- Reset Larry to 1_1 (stage 1 of scenario)
- Reset scenario equipment
  - Attach power cable again (string)
  - Put hatchet, scissors and bomb back in the boxes
  - Restack the alien base

2.3 Implementation

RS Media

Robot Hardware

To implement our interaction, we used a WowWee Robosapiens “RS Media” robot, which we named Larry. The robot is controlled by an embedded processor running Linux which enables new behaviors to be added in software. Larry also has a set of default behaviors that needed to be modified in order for him to only execute the actions given to him. To do this we needed to turn off his vision and his hearing or he would respond to these stimuli with unnecessary actions. Below is a table of constraints we encountered when implementing the interaction:

Aspect	Constraints
Walking	Needs guide to walk straight, cannot turn reliably
Gripping	Can reach out to grasp object, but recognition is too spotty to use
Facial Features	Eyes blink on IR command received, can turn eyes off and on
Speaking	Plays Mp3 files
IR Commands	Max. 10 custom commands, cannot receive commands while performing any action

Figure 3 RSMedia robot constraints

The robot came with a hand-held infrared (IR) remote to control his movements, and to call custom commands. We did not use this remote in our study, but instead replicated the Infrared codes required to call the various onboard commands, which will be described later in more detail.

Robot Software

Mathew S. Howe, a WoWee Robotics enthusiast distributed a table of internal codes for all of RSMedia’s commands, from walking and lifting his right arm to playing and video files. Using this table we were able to create our own BASH shell scripts to control the robot through a series of “stages” needed to complete the scenario. Advancing from one stage to the next required a verbal response from the participant. Whenever a word was uttered by the participant (e.g., “affirmative”), a script associated with that word is run. The script checks a text file on board the robot which records the current stage. If the stage is one that requires the utterance of that word, then another script is called that executes the next stage, advancing Larry. This advancing script then checks what stage Larry is at, and has Larry walk forward if and only if he just finished overcoming the last obstacle and must approach the next one. Then Larry will play the audio corresponding to the next stage. This audio continues the story, letting the participant know what actions he is to perform next.

Figure 4: RSMedia Suite Screenshot

RS Media Suite

Along with the RS Media robot, WowWee provided a custom software suite (see Figure 4) to help generate shell scripts for custom movements and behaviors. The software proved most useful when aligning audio tracks with the robot’s body movements.

Spoken Communication

The communication interface is what allows the participant to talk to Larry and be understood. The biggest difficulty with this interface was the speech recognition program. Larry is only able to understand three words, “affirmative”, “continue”, and “help”. After much experimentation, this small vocabulary was the most reliable. Even with such limited breadth, we believe that using speech was the best way to provide a natural human-like interaction with the robot. One of the experimenters pre-recorded Larry’s utterances using appropriate dramatic and engaging voice presentation. Please refer to Appendix B for the scenario script.

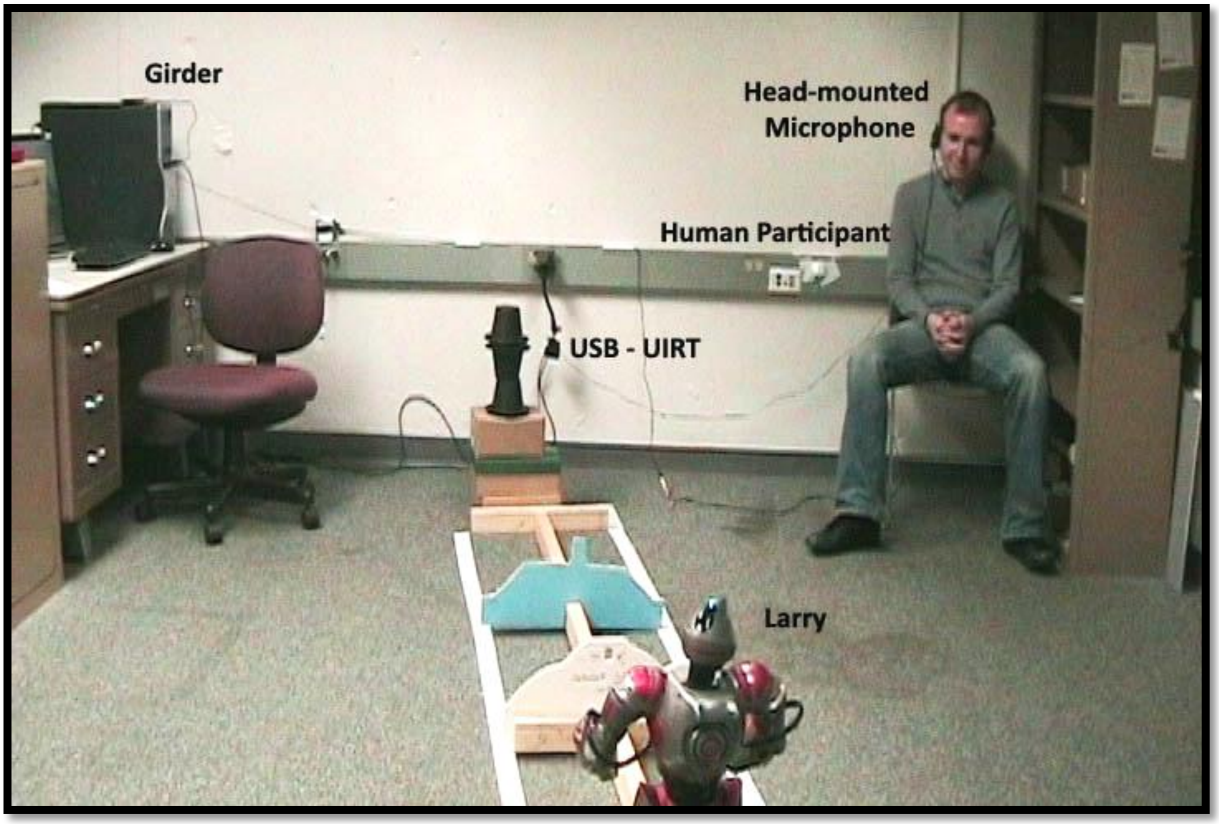


Figure 5: Interaction Setup

Figure 5 illustrates the basic configuration of the experimental equipment. The components include:

- Head-mounted microphone
- Girder (on computer) with plug-ins for:
    - Microsoft Speech Recognizer
    - USB-UIRT
- USB-UIRT, an infrared transmitter and receiver
- RSMedia robot (Larry)

Girder (www.proximis.com) is a home automation software suite which handles voice recognition (via a plug-in for Microsoft Speech Recognizer) and transmits infrared commands. We defined a grammar which allowed Girder to recognize our small vocabulary. The XML grammar definition file is in Appendix D.

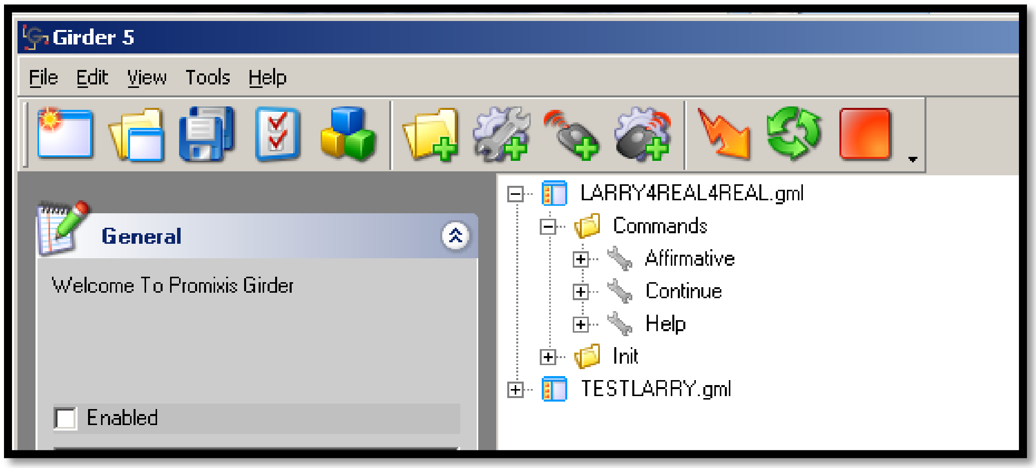


Figure 6: Girder

The Universal Infrared Receiver Transmitter device (USB-UIRT) connects to a computer via a USB port and transmits and receives IR codes. This device is integrated into Girder (interface in Figure 6) via a special plug-in. We used the IR receiving functionality to record custom IR commands from the robot's manual remote control to use in the interaction. After these were recorded, Girder used the USB-UIRT device to transmit IR code to communicate with Larry. Each of the three human command words, as well as each initialization command, was mapped to one of the robot's custom macro commands (using the associated IR code).

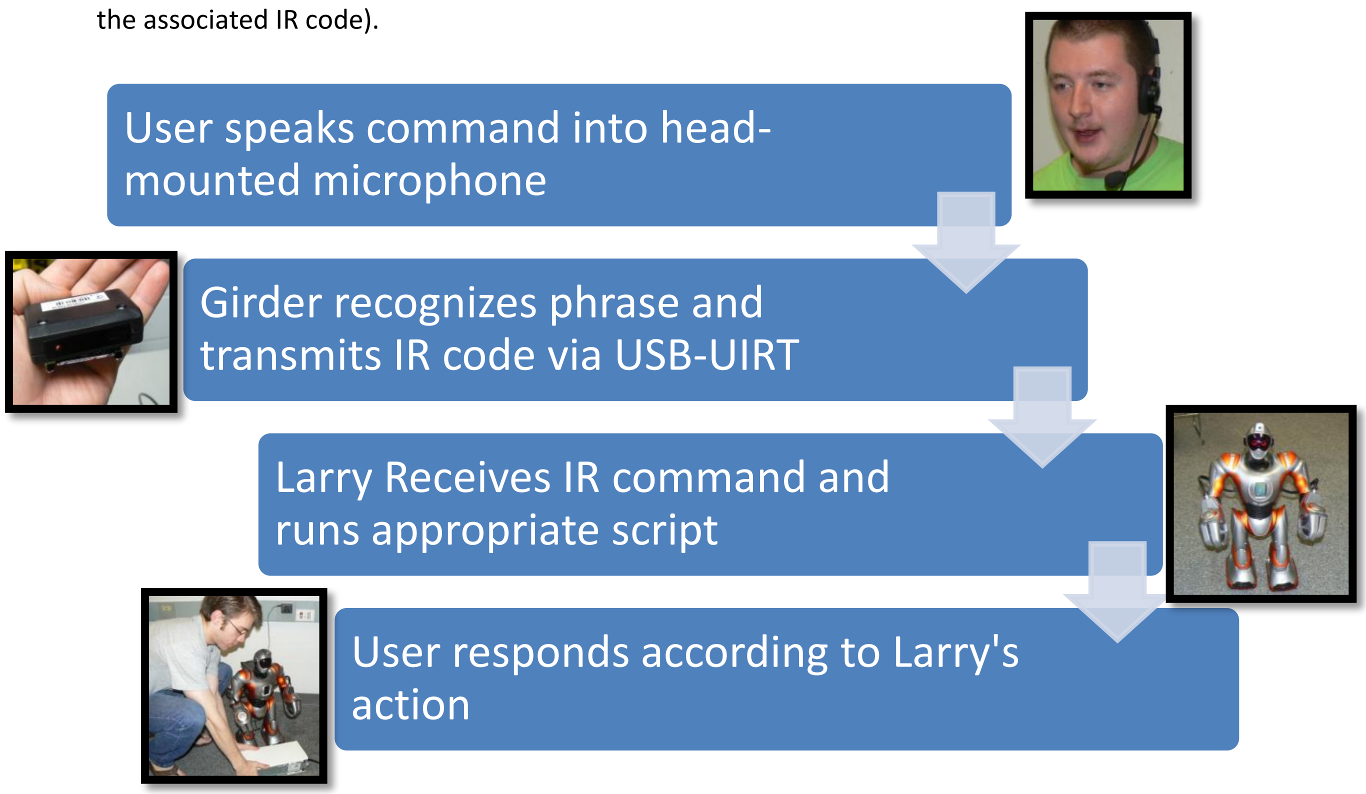


Figure 7: Hardware Communication Interface

An Example Interaction

We used a head-mounted microphone and Microsoft speech recognizer to capture the spoken commands of the study participants. These commands are sent to Girder (home automation software) that transmits the corresponding IR code to Larry. The infrared code is then transmitted through the USB-UIRT device. Once the IR code is sent, Larry will use that information along with where it is currently in the scenario script to determine his next action. Every time a correct IR code is sent to Larry, he will respond with an action, such as moving forward (and pointing if he is in the gesture condition). Next Larry tells the participant what to do next, such as retrieving the hatchet or cutting the power cables. Larry also lets the participant know what they need to say next to continue the scenario. For example, Larry says: "Say 'continue' when the path is clear". A simplified communication flow can be seen in Figure 7.

Study Execution

There were 36 study participants in total over the course of five weeks, though we only used data from 26 of them, because in 10 cases the robot software failed during the interaction and the participant could not complete the interaction. The study was counterbalanced and between-subject so there were 13 people in the gesture condition and 13 in the control condition. Our study population consisted of people ages 18 – 52 in the WPI community. Of the 26 participants who successfully completed the study, 20 were male and 6 were female (5 in the gesture condition). We mostly recruited participants personally although flyers and emails were also used. We allotted 20 minutes of time for each participant, which was ample time to prepare, run the trial, and reset for the next participant.

Chapter 3: Results and Analysis

3.1 Completion Time

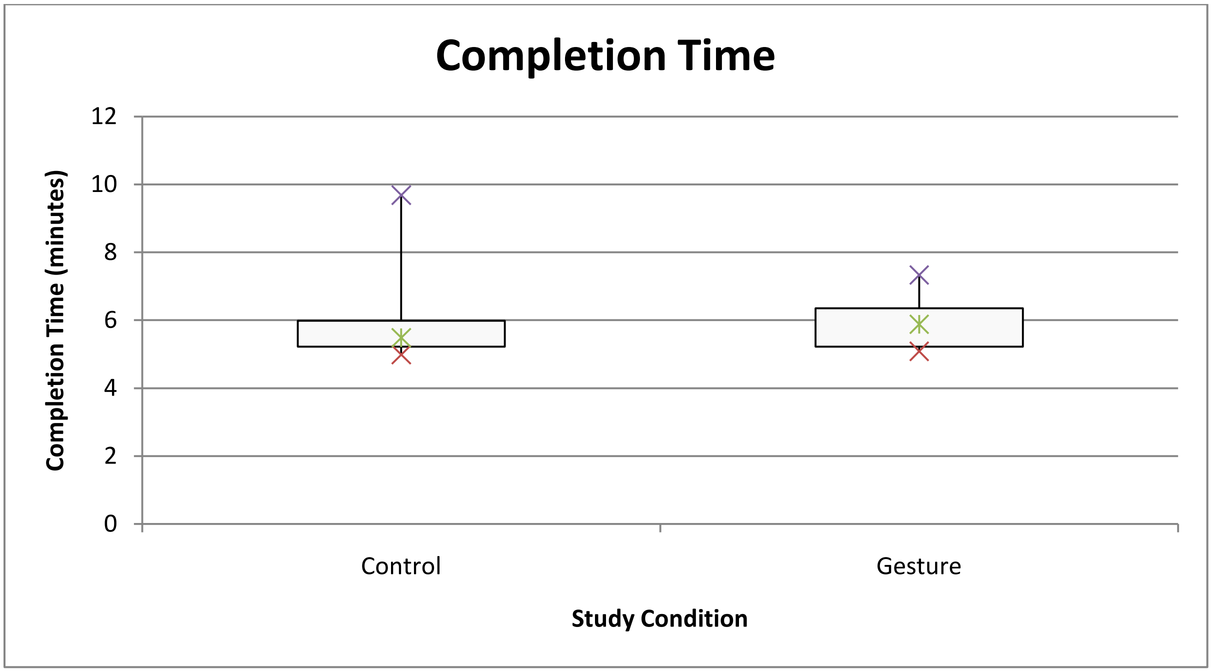


Figure 8: Recorded Completion Time

In Figure 8, we see a box-and-whisker distribution of completion times. The box encapsulates the middle 50% of responses, and the outliers and median are each marked by an X. We found that completion times were very similar between the conditions, and there was no significant difference, with a T-Test p-value of 0.84. It may be noted however that the minimum and maximum times were more extreme in the control condition.

3.2 Questionnaire Data

The questionnaire, completed after participants completed the interaction, included asking the participant to estimate the time it took to complete the scenario and answering a set of 7-point Likert-scale questions. Some questions were "negative" questions (e.g. "I was confused by the scenario objectives") and those values were reversed (1 changed to 7, 2 changed to 6, etc.).

Perceived Completion Time

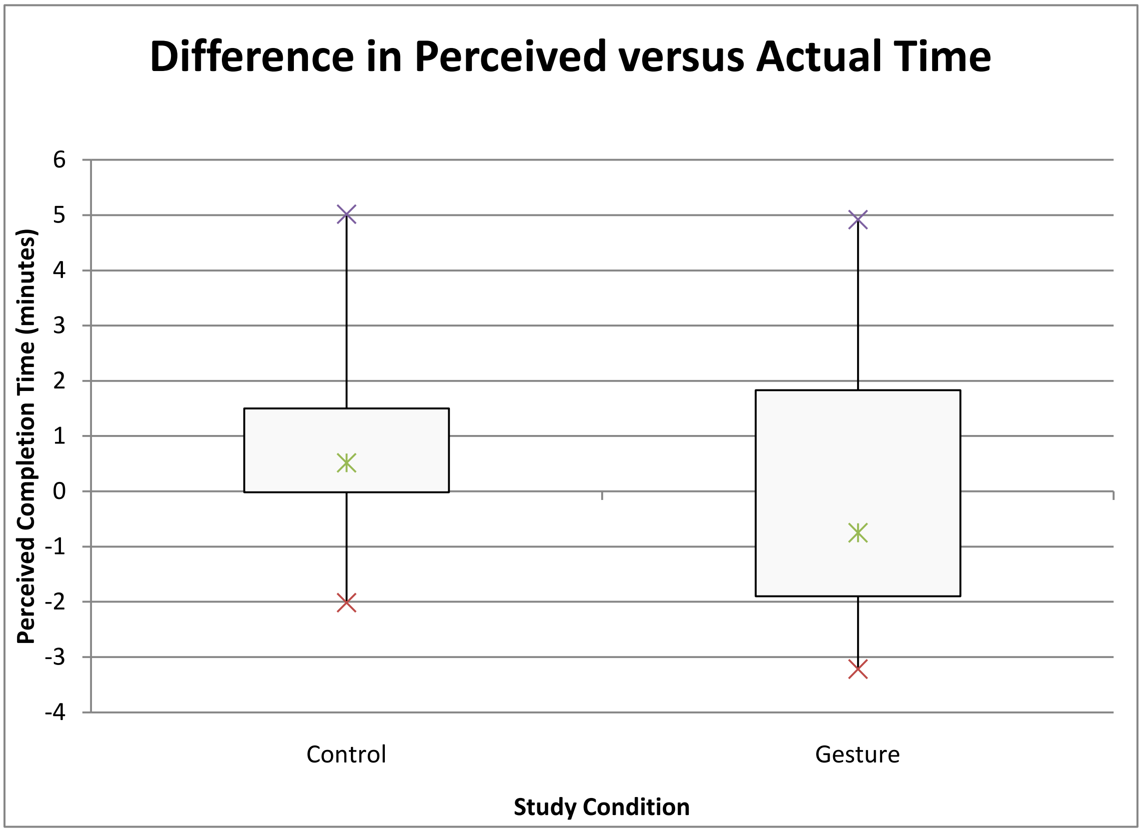


Figure 9: Difference in Perceived vs. Actual Completion Time

Figure 9 shows the difference in the perceived and actual completion time. The data represents "overestimation", so a positive value indicates that the study participant estimated that the interaction took longer than it did. On average, the control condition participants overestimated their time by 0.93 minutes more than the gesture condition. The difference was marginally significant, with a T-Test p-value of 0.26.

Interaction Skills

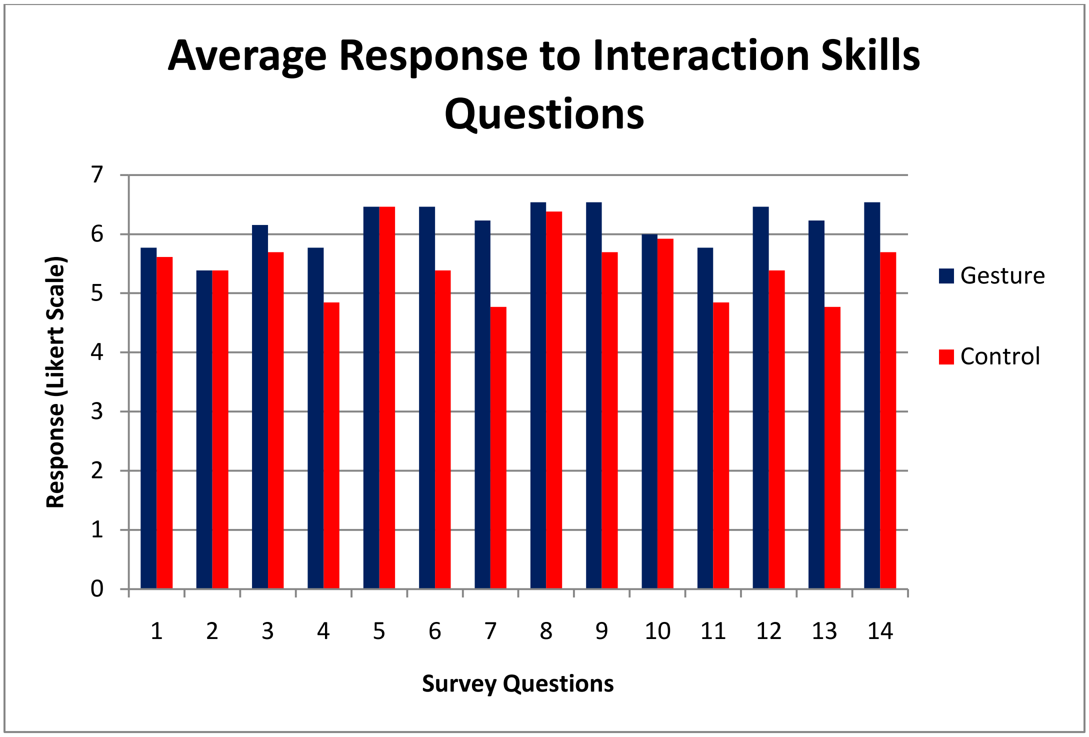


Figure 10: Interaction Skills

Figure 10 shows the difference in average responses for questions regarding interaction skills. We can see that in each case, participants in the gesture condition responded the same or better than in the control condition. The largest difference was observed in question 13: “I knew when Larry did not understand me”.

Mann-Whitney U Test

The p-value obtained by analyzing the grouped data for all questions in this category was less than 0.001, indicating a very significant result.

Collaboration

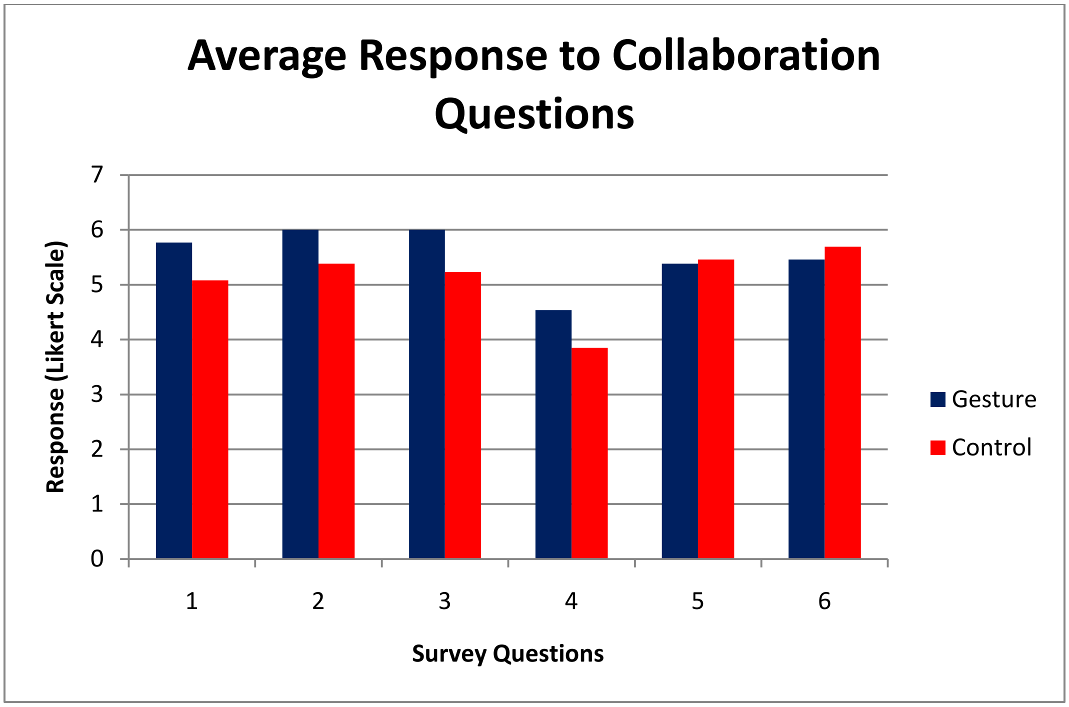


Figure 11: Collaboration

Figure 11 shows the difference in responses for questions relating to Larry's skills in collaboration. These responses were somewhat mixed, with smaller disparity than the Interaction Skills category.

Mann-Whitney U Test

The p-value obtained by analyzing the grouped data for all questions in this category was 0.057, indicating a significant result.

User Experience

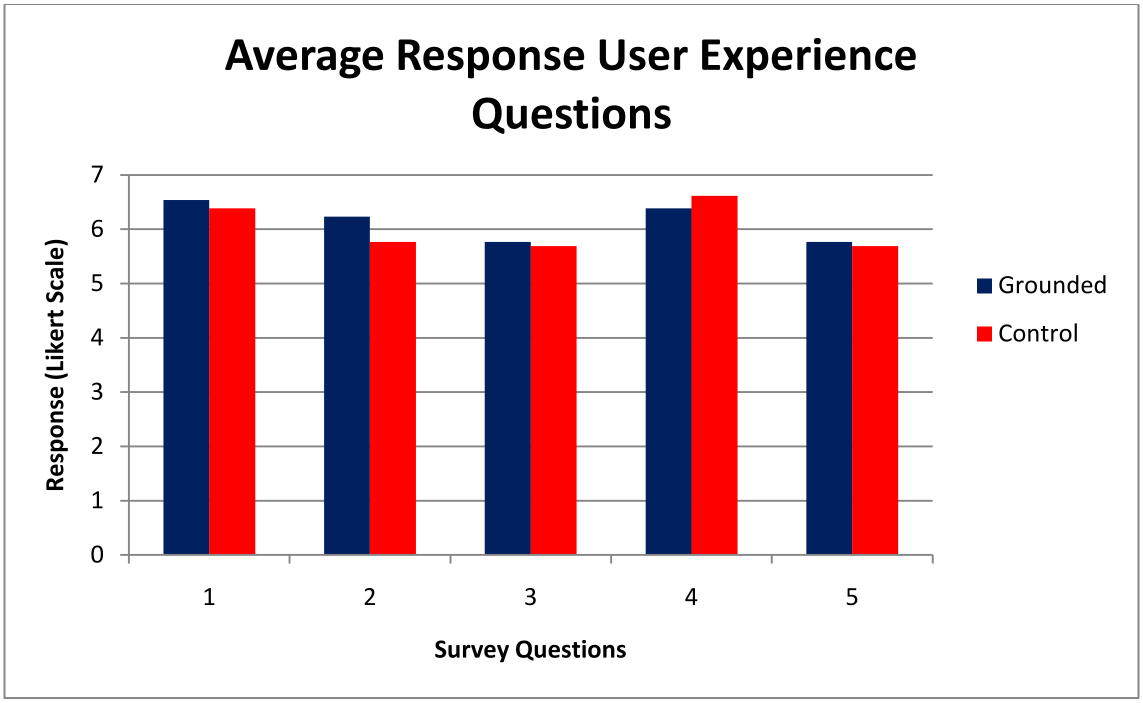


Figure 12: User Experience

Figure 12 shows the difference in responses for questions relating to user experience. There was not much difference in this category.

Mann-Whitney U Test

The p-value obtained by analyzing the grouped data for all questions in this category was 0.706, which did not indicate a significant result.

3.3 Evaluation of Hypotheses

Our conclusions based on data analysis of our 5 hypotheses are presented below:

Participants will think the robot has better interaction skills.
-Confirmed

When Larry turned his head to the participant when speaking, we believe that it helped to grab the participants' attention, making them feel like working with Larry was that working with another human. This helped make the participants feel as though Larry understood them, making Larry appear to have strong interaction and collaboration skills. Below is a picture of a participant involved in the control condition of the experiment. Here the robot is not looking at him, and the participant's attention remains away from Larry.

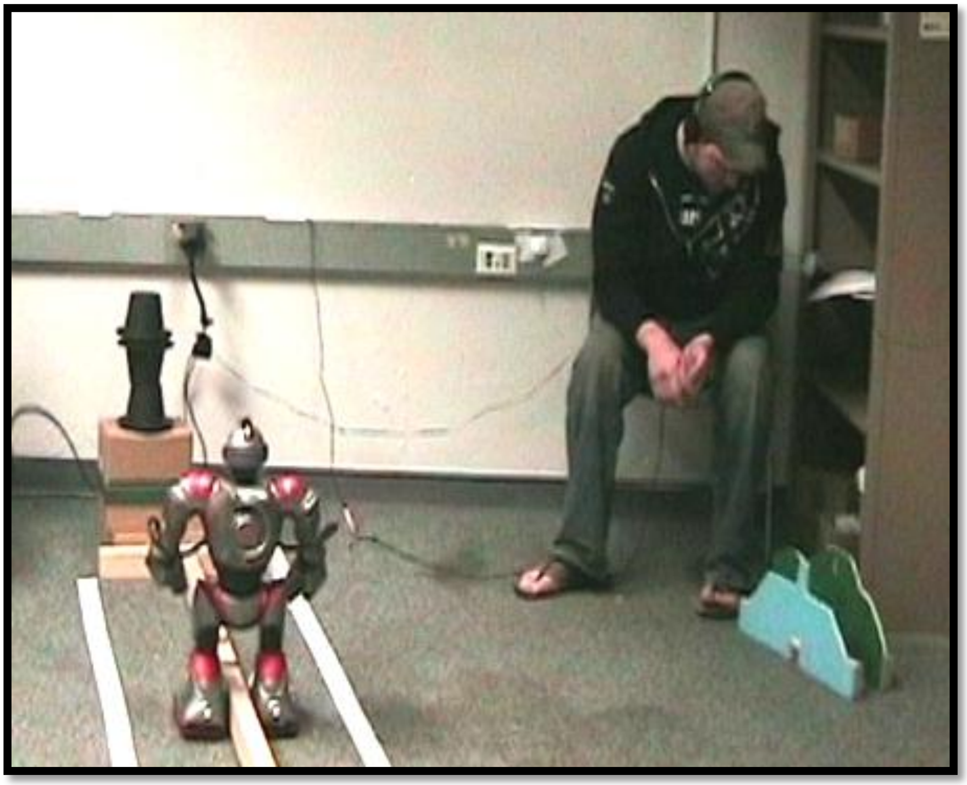


Figure 13 Control Condition

Participants will think the robot is a better collaborator.

-Confirmed

In the gesture condition, participants felt that collaboration with the robot was better than in the control condition. By pointing towards scenario elements, we believe the robot was better able to communicate the location of the objects the participants needed to pick up. More participants immediately turned their attention to objects as a result of this gesture than the participants during the control condition did during the same stage of the scenario. We also believed that this arm gesture contributed to Larry being an effective collaborator because it helped the participants to find scenario objects more easily.

Participants will enjoy the interaction more.

-Not Supported

The third hypothesis was not supported, showing that this experiment did not demonstrate that gestures make the participants feel more comfortable. We observed in video participants appearing more comfortable in the gesture condition, but questionnaire data did not support this. There were also not as many questions in this section, so this is an area which could be further explored.

Completing the goal of the interaction will take less time

-Not Supported

We found no evidence that scenario completion time was in any way improved.

Participants will perceive that completing the goal of the interaction took less time

-Not Supported

We found marginal support for Hypothesis 5 although there was not quite enough data to conclude that the hypothesis was satisfied. If similar results were achieved on a larger study group, this could become significant.

Chapter 4: Conclusion and Recommendations

In order to improve human-robot interaction, it is important to understand human-human interaction and integrate these same principles in robots. When robots are needed to help humans with a task, both parties need to understand each other and carry out the tasks required to satisfy them both. In human-computer interaction (HCI) feedback is extremely important for any kind of interface [Cooper 1995, Fry & Smith 1975]. In a way, interfaces primarily are trying to communicate information to you in the best way possible, while making it easier to convey information to them so they can perform certain tasks. Clicking on a song and hitting the play button on media software is an example of this. Robots are computers with the likeness of humans, and they need a much more developed system to communicate with humans because they are a man-made representation of humans.

Our study demonstrates that the implementation of even very simple gestures (such as pointing to topics of interest and gaze) can improve the experience of human users interacting with robots. Developing this principle will lead to a better human-robot interface design.

We found it interesting that the time to completion was virtually identical between the two conditions on average. This is surprising because in general you would think that if a robot has better collaboration and interaction skills, then completing tasks should be faster. In this case, the addition of gestures could have added some overhead to the time of the interaction and any improvements only served to recoup those losses. We still believe that stronger collaboration and interactive skills will in general lead to completing tasks more quickly. One of the reasons that the difference in completion time may not be significant between the two conditions could be that there are too many other factors contributing to this measure. An experiment focusing more on this measure would be more helpful in confirming that collaboration and interaction skills in robots do help contribute to task completion time. It should also be noted that the difference in *perceived* time to completion was marginally significant, showing that people believed the interaction was shorter when they were engaged with a robot that used gestures. This supports the idea that people feel more at ease when interacting with a robotic system with human-like gestures.

Developing quantifiable data is a challenge when studying the effects of gesture in human-robot interaction. The measures used in this study have been useful in finding a connection between gestures and collaboration. Better understanding the features behind human-human interaction will help find other useful measures (Sidner et al. 2005).

We would like to conclude by pointing out two promising directions to extend and continue the kind of work we started. First, there may be a lot of useful information in our video data beyond just completion time. For example, using the video data we collected, it might be possible to find a method for analyzing the error rate of the participants. Data that might be compared between the gesture and control conditions include: (1) the number of times a participant asks for help, (2) how often a participant has to repeat himself/herself before Larry responds, and (3) how often a participant is confused about the location of an item.

Second, it would be interesting to study the effects of other types of non-verbal or related behaviors on human-robot collaboration. There are a number of different behaviors seen between humans that support collaboration. In this study we focused on the effect of two types of gestures: head turns and pointing, and the effects they had on human-robot interaction. Another type of behavior that could be looked at in the future is the effect of *grounding* on human-robot interaction. In human-human interaction, grounding refers to the set of information that is both known by all participants and known to be known by all participants. Conversations develop their grounding over time, constantly contributing information, and confirming that the other person understood what they were telling them [Clark & Schaefer 1989]. According to Billard & Dautenhahn, in a situation where a robot needs to communicate with a human, both agents need to share an understanding of one language [1997]. Aside from basic verbal statements, there are a number of subtle behaviors that are used to contribute to grounding. Head nods, facial expressions and verbal utterances all help to ground a conversation. Studying the effects of these behaviors on human-robot interaction will help to improve these interactions in the future.

Appendices

Appendix A: Raw Questionnaire Data

Interaction Skills

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Control	5	6	4	7	7	7	7	7	7	7	7	7	7	7
Control	6	6	6	5	7	7	3	6	7	7	6	7	3	7
Control	6	6	4	6	7	7	5	7	7	5	6	7	5	7
Control	6	7	6	4	6	5	3	7	5	3	1	5	3	5
Control	6	5	3	6	7	2	2	6	5	7	4	2	2	5
Control	5	4	6	6	7	3	4	7	3	6	5	3	4	3
Control	5	3	6	5	4	6	5	5	6	6	4	6	5	6
Control	7	6	7	7	7	7	7	7	5	7	7	7	7	5
Control	5	5	7	5	7	5	7	6	7	7	3	5	7	7
Control	5	4	5	2	5	4	6	6	5	3	2	4	6	5
Control	5	5	7	1	7	3	2	6	3	5	5	3	2	3
Control	6	6	7	5	7	7	5	6	7	7	7	7	5	7
Control	6	7	6	4	6	7	6	7	7	7	6	7	6	7
Gesture	6	6	6	5	7	7	6	7	6	7	6	7	6	6
Gesture	5	5	5	3	7	7	5	6	6	6	6	7	5	6
Gesture	6	6	6	7	6	7	7	7	7	7	7	7	7	7
Gesture	3	3	5	6	4	5	6	6	7	3	3	5	6	7
Gesture	6	5	6	5	6	7	5	6	6	7	6	7	5	6
Gesture	6	7	7	6	7	6	7	7	7	6	5	6	7	7
Gesture	6	5	7	6	7	7	7	7	7	7	6	7	7	7
Gesture	7	6	7	7	7	7	7	7	7	7	6	7	7	7
Gesture	6	6	7	6	7	6	6	7	6	7	6	6	6	6
Gesture	7	7	7	6	7	7	6	7	7	3	5	7	6	7
Gesture	6	4	6	5	6	5	6	6	6	7	7	5	6	6
Gesture	6	6	7	7	7	7	7	7	7	4	5	7	7	7
Gesture	5	4	4	6	6	6	6	5	6	7	7	6	6	6

1. I found Larry easy to interact with.
2. Larry responded well to me.
3. I feel that Larry understood me.
4. I understood Larry's body language.
5. Larry did things I didn't want him to do.*
6. I always knew when Larry understood what I told him to do.
7. I knew when Larry did not understand me.
8. I enjoyed the interaction.
9. I could tell when Larry understood me.

10. I was confused by the scenario objectives.*
11. I always knew what to do during the interaction.
12. I always knew when Larry understood what I told him to do.
13. I knew when Larry did not understand me.
14. I could tell when Larry understood me.

Collaboration

	1	2	3	4	5	6
Control	5	5	6	5	6	5
Control	4	5	3	3	5	5
Control	4	4	7	2	4	5
Control	5	6	4	5	5	5
Control	5	5	4	3	6	7
Control	7	7	6	5	6	7
Control	5	5	6	3	5	2
Control	7	7	7	7	7	7
Control	4	4	5	3	6	6
Control	4	4	5	5	5	5
Control	4	5	3	1	5	7
Control	6	7	6	5	5	6
Control	6	6	6	3	6	7
Gesture	6	6	5	5	6	6
Gesture	6	6	5	5	5	5
Gesture	6	7	7	6	5	6
Gesture	3	3	6	3	4	4
Gesture	5	5	6	4	5	3
Gesture	7	7	7	7	7	7
Gesture	5	6	6	4	6	7
Gesture	7	7	5	3	4	7
Gesture	7	7	7	5	7	5
Gesture	7	7	6	5	6	7
Gesture	5	6	5	2	4	4
Gesture	7	7	7	7	7	6
Gesture	4	4	6	3	4	4

1. Larry collaborated well with me.
2. Larry and I worked well together.
3. I felt engaged by Larry
4. Working with Larry was like working with another person.
5. I thought Larry was a good partner.
6. I could depend on Larry to work correctly every time.

User Experience

	1	2	3	4	5
Control	7	7	7	7	7
Control	6	6	6	6	6
Control	7	7	5	7	5
Control	7	6	4	5	4
Control	6	6	7	7	7
Control	7	7	6	7	6
Control	5	5	3	6	3
Control	7	7	7	7	7
Control	6	5	5	7	5
Control	6	5	5	6	5
Control	6	4	5	7	5
Control	6	6	7	7	7
Control	7	4	7	7	7
Gesture	7	7	6	6	6
Gesture	6	5	6	6	6
Gesture	7	7	6	7	6
Gesture	6	6	3	6	3
Gesture	6	5	5	6	5
Gesture	7	7	7	7	7
Gesture	7	5	6	6	6
Gesture	7	7	7	7	7
Gesture	7	7	6	7	6
Gesture	7	7	6	6	6
Gesture	6	6	6	7	6
Gesture	7	7	7	7	7
Gesture	5	5	4	5	4

1. I enjoyed the interaction.
2. I felt involved in the scenario.
3. The task was accomplished quickly.
4. I felt comfortable during the task.
5. The task was accomplished quickly.

Appendix B: Scenario Script: "Gauntlet"

Setup:

- Several Stages
    1. Removing the foliage
    2. Shield Generator
    3. Bomb assembly (robot instructs human)
    4. Pick up bomb
    5. Use bomb to destroy base

A play through: (L=Larry, H=Human)

~~~ Introduction ~~~

1.1

L: Greetings! My sensors indicate that you are a trustworthy humanoid. My name is Larry and I have some problems at hand. There is a race here called the Zorgs who are trying to destroy our once hidden home planet of 404Error! They have setup base not far from here. I need to destroy it, but I cannot make it alone. Are you willing to help me?!

H: Affirmative

~~~ Removing the foliage ~~~

Figure 14 Forest

2.1

L: Their base is just beyond this forest, and I need your help to clear the way. To do this, you must first acquire the greatest of all hatchets, for time is of the essence! Quick, open box number one. Give me the Affirmative when you have acquired it!

H: <Goes to cabinet and takes out the 'Hatchet of Arboreal Submission'. Walks back to place> Affirmative.

2.2

L: Good! Now use the hatchet to uproot the trees! Tell me to Continue when the path is clear!

H: <Lifts the "trees" out of the way> Continue

~~~ Shield Generator ~~~

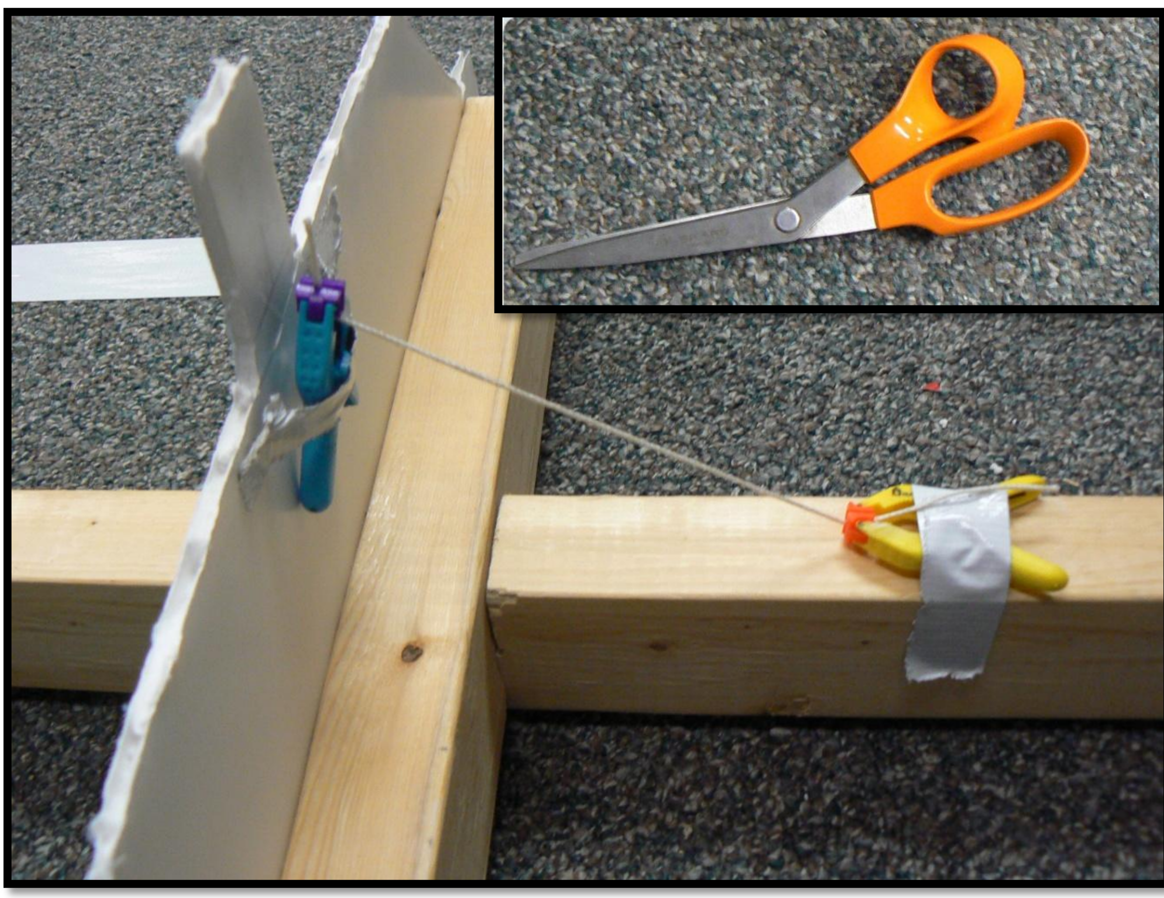


Figure 15 Shield Generator

3.1
~~~

L: <Moves forward and stops at weight> I've reached the shield generator. Quick, look in box number two! Use the scissors to cut the power cable to the generator. We'll have this cleared in no time. Give me the affirmative when it's cut!

H: Affirmative

3.2

L: Now remove the device and tell me to Continue when the path is cleared.

H: <Clears the generator> Continue

L: <Moves forward and stops at weight>

~~~ Infiltrate Base~~~

4.1

L: Now make sure to be quiet, because we are infiltrating their base now. Find the bomb in box number three! Affirm me when you have obtained it!

H: Affirmative

~~~ Pickup Bomb ~~~

Figure 16 Bomb

5.1

L: Now I need you to take the bomb and place it at the base of the tower. Tell me to continue when it's in place.

H: Continue

5.2

L: The time has come. I am ready and not nervous at all. Do you want to know why?

H: Affirmative

5.3

L: It's because I have nerves of steel, quite literally. Haha! Ok, now tell me to Continue when you're ready.

H: Continue.

~~~ Exploding the Building~~~

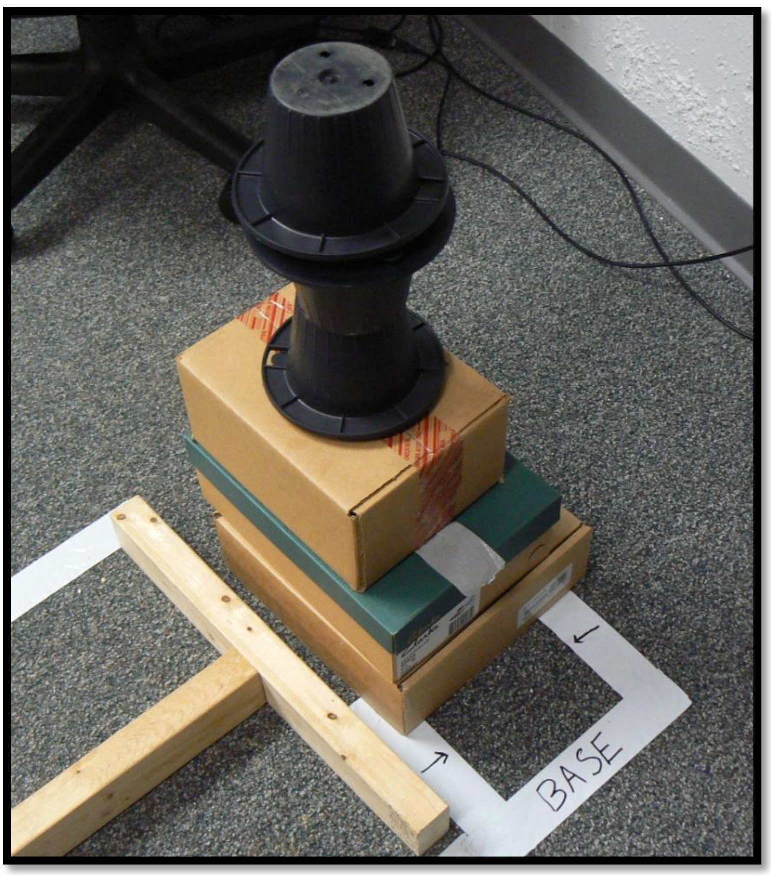


Figure 17 Alien Base

6.1
~~~

L: <Moves forward and stops at weight in front of base> Arming bomb… Ok it is ready now! Tell me to Continue when you're safely clear of the building!

H: Continue

~~~ End ~~~

7.1

L: ***EXPLOSION*** We did it! My home planet is much safer thanks to you, stranger! Please head to the computer by the door and fill out the exit survey, then help yourself to refreshments outside! Thanks again, and goodbye. <Larry shuts down into sleep mode>
~~~

Appendix C: Informed Consent

Informed Consent Agreement for Participation in a Research Study

Investigator: Charles Rich, Joel Sutherland, Kevin O'Brien

Contact Information:

Professor Charles Rich
Computer Science Department
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609
Tel. 508-831-5945

Title of Research Study: Design of a Spoken-Language Interface

Introduction:
You are being asked to participate in a research study. Before you agree, however, you must be fully informed about the purpose of the study, the procedures to be followed, and any benefits, risks or discomfort that you may experience as a result of your participation. This form presents information about the study so that you may make a fully informed decision regarding your participation.

Purpose of the study:
This study is to determine if implementing human-like collaboration techniques in robotic systems improves the performance and experience of working in a human-robot team.

Procedures to be followed:
As a subject in this study you will be asked to participate in a collaborative task with a robot. You will be given a headset with which you will communicate with the robot, and you will need to assist the robot physically at several steps.

Risks to study participants:
You will be asked to sit in a chair and wear a headset to communicate with the robot. Also, you will need to stand and lift several objects weighing not more than ten pounds.

Benefits to research participants and others:
This study will help to advance the development of robot interface design. Participants will be able to interact with a unique robot personality and experience a story-based interaction.

Record keeping and confidentiality:
Records of your participation in this study will be held confidential so far as permitted by law. However, the study investigators, the sponsor or it's designee and, under certain circumstances, the Worcester Polytechnic Institute Institutional Review Board (WPI IRB) will be able to inspect and have access to confidential data that identify you by name. Any publication or presentation of the data will not identify you. We will keep video recordings of your participation in the

study, as well as your written answers to survey questions after completion of the study. These records will be kept in a secure office and never associated with you by name.

Compensation or treatment in the event of injury:
In the unlikely event of physical injury resulting from participation in the research, you understand that medical treatment may be available from WPI, including first aid emergency care, and that your insurance carrier may be billed for the cost of such treatment. No compensation for medical care can be provided by WPI. You further understand that making such medical care available, or providing it, does not imply that such injury is the fault of the investigators. You do not give up any of your legal rights by signing this statement.

For more information about this research or about the rights of research participants, or in case of research-related injury, contact:

Professor Charles Rich
Computer Science Department
Tel. 508-831-5945
Email: rich@wpi.edu

Professor Kent Rissmiller
IRB Chair
Tel. 508-831-5019
Email: kjr@wpi.edu

Michael J. Curley
University Compliance Officer
Tel. 508-831-6919
Email: mjcurley@wpi.edu

Your participation in this research is voluntary. Your refusal to participate will not result in any penalty to you or any loss of benefits to which you may otherwise be entitled. You may decide to stop participating in the research at any time without penalty or loss of other benefits. The project investigators retain the right to cancel or postpone the experimental procedures at any time they see fit.

By signing below, you acknowledge that you have been informed about and consent to be a participant in the study described above. Make sure that your questions are answered to your satisfaction before signing. You are entitled to retain a copy of this consent agreement.

By selecting yes directly below this paragraph, you voluntarily give permission for short excerpts and snapshots from this video to be used for public scientific purposes, such as in published papers, books, seminars and conference presentations, and to show to other researchers. You agree not to expect any compensation for this permission and understand that no identifying information, such as my name or email address, will be associated with this use.

☐ Yes ☐ No

____________________________ Date: ____________________

Study Participant Signature

____

Study Participant Name (Please print)

______________________________________ Date: ____________________

Signature of Person who explained this study

Appendix D: XML Grammar Used in Girder

```
<GRAMMAR LANGID="409"><!-- English -->

	<RULE NAME="AFFIRMATIVE" TOPLEVEL="ACTIVE">
		<P>Affirmative</P>
	</RULE>

	<RULE NAME="CONTINUE" TOPLEVEL="ACTIVE">
		<P>Continue</P>
	</RULE>

	<RULE NAME="HELP" TOPLEVEL="ACTIVE">
		<P>Help</P>
	</RULE>

</GRAMMAR>
```

Appendix E: Command Sheet

Affirmative

Continue

Help

** NOTE: Count to 5 before responding to Larry, There is a slight delay before he understands you.

Appendix F: Questionnaire

Likert Scale

1. I found Larry easy to interact with.
2. Larry responded well to me.
3. I feel that Larry understood me.
4. I understood Larry's body language.
5. Larry did things I didn't want him to do.
6. Larry collaborated well with me.
7. Larry and I worked well together.
8. I felt engaged by Larry
9. Working with Larry was like working with another person.
10. I thought Larry was a good partner.
11. I always knew what to do during the interaction.
12. The task was accomplished quickly.
13. I was confused by the scenario objectives.
14. I could depend on Larry to work correctly every time.
15. I felt involved in the scenario.
16. I always knew when Larry understood what I told him to do.
17. I knew when Larry did not understand me.
18. I enjoyed the interaction.
19. I could tell when Larry understood me.
20. The task was accomplished quickly.
21. I felt comfortable during the task.

Free Response

1. Estimate how long it took you to complete the interaction.

References

Argyle, M., M. Cook, and M. Argyle. 1976. *Gaze and mutual gaze*Cambridge University Press Cambridge.

Caproni, V., D. Levine, E. O'Neal, R. McDonald, and G. Garwood. 1977. Seating position, instructor's eye contact availability, and student participation in a small seminar. *Journal of Social Psychology* 103, (2): 315-6.

Clark, H. H. 1996. Using language. *Computational Linguistics* 23, (4).

Clark, H. H., and E. F. Schaefer. 1989. Contributing to discourse. *Cognitive Science* 13, (2): 259-94.

Cooper, A. 1995. *About face: The essentials of user interface design*John Wiley & Sons, Inc. New York, NY, USA.

Duncan, S. 1972. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology* 23, (2): 283-92.

Fry, R., and G. F. Smith. 1975. The effects of feedback and eye contact on performance of a digit-coding task. *Journal of Social Psychology* 96, (1): 145-6.

Goodwin, C. 1986. Gestures as a resource for the organization of mutual orientation. *Semiotica* 62, (1-2): 29-50.

Jarvella, RJ, and W. Klein. 1982. *Speech, place, and action: Studies of deixis and related topics* John Wiley & Sons.

Jellison, J. M., and W. J. Ickes. 1974. The power of the glance: Desire to see and be seen in cooperative and competitive situations. *Journal of Experimental Social Psychology* 10, (5): 444-50.

Kendon, A. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26, : 22-63.

Nakano, Y. I., G. Reinstein, T. Stocky, and J. Cassell. 2003. Towards a model of face-to-face grounding. Paper presented at Proceedings of the 41st Annual Meeting on Association for Computational Linguistics, .

Sidner, C. L., C. Lee, C. D. Kidd, N. Lesh, and C. Rich. 2005. Explorations in engagement for humans and robots. *Artificial Intelligence* 166, (1-2): 140-64.