Exploring Adaptive Dialogue Based on a Robot's Awareness of Human Gaze and Task Progress

Cristen Torrey, Aaron Powers, Susan R. Fussell, Sara Kiesler
Human Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
{ctorrey, apowers, sfussell, kiesler}@cs.cmu.edu

ABSTRACT
When a robot provides direction—as a guide, an assistant, or as an instructor—the robot may have to interact with people of different backgrounds and skill sets. Different people require information adapted to their level of understanding. In this paper, we explore the use of two simple forms of awareness that a robot might use to infer that a person needs further verbal elaboration during a tool selection task. First, the robot could use an eye tracker for inferring whether the person is looking at the robot and thus in need of further elaboration. Second, the robot could monitor delays in the individual's task progress, indicating that he or she could use further elaboration. We investigated the effects of these two types of awareness on performance time, selection mistakes, and the number of questions people asked the robot. We did not observe any obvious benefits of our gaze awareness manipulation. Awareness of task delays did reduce the number of questions participants' asked compared to our control condition but did not significantly reduce the number of selection mistakes. The mixed results of our investigation suggest that more research is necessary before we can understand how awareness of gaze and awareness of task delay can be successfully implemented in human-robot dialogue.

Categories and Subject Descriptors
H.1.2 [Models and Principles]: User/Machine Systems – *Human factors, Software psychology*. H.5.2 [Information Interfaces and Presentation]: User Interfaces – *Evaluation/methodology, Theory and methods*.

General Terms
Design, Experimentation, Human Factors, Theory.

Keywords
Human-robot interaction, human-robot dialogue, adaptive dialogue, social robots.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
HRI'07, March 10-12, 2007, Arlington, Virginia, USA.
Copyright 2007 ACM 978-1-59593-617-2/07/0003…$5.00.

1. INTRODUCTION
Robots that provide instruction, directions, or other types of information may interact with people of various backgrounds and levels of expertise. Different people require different amounts of detail in their directions, depending on their understanding of the task. For example, a cooking expert would understand a robot's direction to "poach" an egg whereas someone less familiar with cooking methods would require further elaboration on the specific steps involved in poaching an egg. Our previous work with a robot chef (see Figure 1) shows that adapting to the particular knowledge requirements of experts and novices improves performance and results in better evaluations of the conversation and the robot [23].

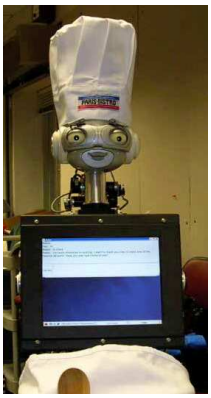

Figure 1. Pearl as a robot chef

Recognizing that robots should engage in some form of adaptation, it remains unclear how precise this adaptation needs to be and how it should be implemented. In our previous experiment, which used a cooking tool selection task, we inferred our participants' knowledge of cooking tools by quizzing them about several cooking methods. By gauging their knowledge of cooking methods, we could infer their likely expertise on cooking tools. People who know how to poach an egg are also likely to know quite a bit about cooking tools.

This general approach could be expanded into a more complete user model for a robot, specific to the cooking domain. When a person demonstrated knowledge of the word "poach" while conversing with the robot, the robot would calculate the likelihood that the person has other cooking knowledge based on the distribution of such knowledge in the population. In that case, the robot could be fairly confident that the person also knew how to sauté, for example, and it would not need to elaborate on such a

direction. A user model, as just described, would require a model of how domain knowledge is distributed in the population, most likely obtained through surveys. This distribution would need to be created for different domains and for different groups of people with whom the robot might interact; thus user modeling would be a time-intensive way to create adaptive robots.

An alternative or supplementary approach for a robot to take would be to elicit and respond to cues given by the person that he or she did not understand the robot's instruction. The robot might ask the person to poach an egg without knowing anything about whether he or she knows how to do so. If made aware of facial expressions or gaze direction, the robot could infer from these signs that the person is confused. Made aware of what expected task progress should be, if the robot noticed the person had picked up a whisk rather than an egg poacher, the robot would elaborate on the previous instruction. Pauses in task progress, incorrect actions, or simply not attending to the task could be useful cues for confirming that the robot's instruction is inadequate, whatever an individual's skill level. By acting on these signals, the robot may be able to elaborate only when it is necessary.

This work explores the usefulness of interpreting human gaze and task progress as signals that a robot's instruction is inadequate and should be elaborated. In this paper, we report the results of a technology trial with four conditions. In the first condition, "Questions Only," the robot initiated no elaborations at all but did respond to participants' questions. In a second condition, "Gaze Added," the robot also had an awareness of the participant's gaze activity and elaborated when the participant was assumed to be looking back at the robot. In the "Delay Added" condition, the robot also had awareness of the participants' task progress. If the participant delayed making a choice in the selection task, the robot elaborated on its previous instruction. In the final condition, "Immediate Added," the robot immediately gave elaboration on all the tools without waiting for a delay in task progress.

The primary goal of this investigation was to understand whether these strategies would be viable approaches to the problem of inferring whether a listener needs further elaboration when a robot is giving direction. A secondary goal was to investigate specific effects of a robot's responsiveness to delays in task progress and to human gaze on participants' performance, communication, and subjective evaluations of the robot and their interaction with it.

2. THEORY & HYPOTHESES

Our work makes use of the literature on common ground [3][22]. Common ground refers to the beliefs and statements that are mutually agreed upon over the course of a conversation. The grounding process is the interactive process of offering statements and accepting them as understood, thus establishing them as common ground. The grounding process is supported by the ability to make inferences about one's listener based on age, gender or group membership [7][10]. The grounding process also benefits from the cycle of listener feedback and utterance repair. Speakers do not need to receive explicit requests from listeners in order to initiate a repair; speakers prefer to repair their own messages [17]. Speakers then use this feedback from the listener to adapt future turns. The principle of "least collaborative effort" suggests that speakers' messages take listeners' needs into account because it avoids the additional effort required by both parties of repairing a confusing or incomplete message [5].

2.1 Return of Gaze Signals Lack of Grounding

Another mechanism that allows speakers to interpret whether their message has been understood is the listener's head and face. If available, facial expressions of confusion are useful, though sometimes difficult to interpret. Research on gaze has demonstrated several important functions, including indicating attention and liking [1][11]. Gaze can also serve as a turn-taking cue [e.g., [6]]; before taking a turn to speak, individuals will seek eye contact with the listener. The direction of the listener's head and gaze are also important signals of attention and comprehension. Nakano et. al. found that speakers watch their listeners' gaze when they refer to a new object [15]. If the listener looks at the object, it has become grounded. If the listener continues to look at the speaker, then further elaboration is required. In this study, we interpret participants' gaze direction in a similar way. We assume that continued gaze toward the tools is an indication that participants are looking for the tool and do not need more information. When a participant turns away from the tools without making a selection, the robot assumes the participant wants to ask a question and further elaborates on the tool. The question-asking process can be time consuming and those without expertise are often not good at formulating productive questions. We predict that responsiveness to participants' gaze in this manner will provide the elaborations necessary to complete the task with fewer mistakes, asking fewer questions of the robot, and ultimately requiring less time.

2.2 Delays in Task Progress Signal Lack of Grounding

When a speaker is unsure whether a message will be understood by a listener, he or she can attempt to mitigate the listener's efforts in formulating a question by paying close attention to the listener's response [5]. If the listener does not confirm that the message is understood, either explicitly in words or by taking an appropriate action, the speaker can volunteer further information in order to reduce the overall communicative effort involved. Thus, studies of communication that contrast pairs with and without information about their partners' task activity demonstrate the importance of visual information in the grounding process [8][12][4]. Participants with awareness of their partners' task activity, use that awareness to monitor progress and decide if further elaboration is necessary. Based on the literature demonstrating the importance of visual information, we predict that a simple form of this awareness will improve the quality of human-robot dialogue.

We use a simple delay in task progress as an indicator that the robot's instruction was not adequate for the listener. The robot is aware when a tool has not been selected in a specified amount of time. The robot then elaborates on its description of that tool, providing additional information that should enable the listener to find it. Although the individual has not yet asked for help, we assume the delay represents uncertainty in the choice, so we provide elaboration in order to prevent error.

2.3 Interpreting Feedback from Experts & Novices

In our previous investigation of appropriate elaboration, we confirmed that experts and novices have different information

requirements when receiving direction from a robot about a cooking task [23]. The current study explored responsiveness to task progress and human gaze because we inferred from common ground theory that these two forms of awareness would benefit both novices and experts. If so, the robot would not be required to classify the listener's expertise.

In the previous study, novices benefited from, and appreciated, full instruction whereas experts generally needed only names of cooking tools to identify them correctly. In the current study, the robot never gave full instruction unless prompted by participants' repeated questions. Instead, the robot gave only partial elaboration as needed for the participant to make a choice. The forms of awareness and robot response we studied in this exploration should not advantage some listeners over others. It is possible, nonetheless, that novices, as compared with experts, would rather have immediate elaboration than wait for the robot to perceive they are confused. In the event this might be the case, we included measures of participants' cooking expertise and the same measures of the robot's communication competence and effectiveness as in the previous study.

3. METHOD

We used a cooking tool selection task identical to that described in [23]. The robot asked each participant to select ten cooking tools needed to make a crème brûlée dessert. Participants selected the tool by clicking on the correct picture on a computer monitor. Each of the ten tools was displayed separately alongside five incorrect tools (see Figure 2). The robot verbally led the participant through the task, requested each of the tools in turn, and answered the participants' questions. Participants could ask the robot as many questions as they wished before selecting a tool. The robot informed participants if they made an incorrect selection. Participants made as many attempts as necessary to select all ten tools correctly.

Figure 2. The tool selection task

3.1 Participants

Sixty-six students and staff members with no prior participation in our experiments were recruited from Carnegie Mellon University. They were each paid $10 for their participation.

3.2 Trial Conditions

Our exploration of the effects of a robot's responsiveness to human gaze and task delay included four trial conditions (see Table 1). In the baseline condition, the robot responded only to questions articulated by the participant. The next two conditions added increasing levels of awareness to the robot's capabilities with the addition of responsiveness to gaze and responsiveness to a delay in task progress. In the fourth and final condition, we isolated the effect of the delay by adding a condition which offered elaborations immediately, as each tool was introduced. We describe these conditions in more detail below.

Table 1. The robot offers elaboration differently in each of the four trial conditions

Condition	After Questions	After Gaze	After Delay	Immediately
Questions Only	✓			
Gaze Added	✓	✓		
Delay Added	✓	✓	✓	
Immediate Added	✓	✓		✓

3.2.1 Questions Only Condition

In the Questions Only Condition, the robot conversed with the participant, only offering additional information if the participant verbally requested it. For example, the robot introduced the paring knife by saying "Next you want a sharp paring knife. Find the paring knife." Participants who did not know which selection to make asked specific questions or told the robot they needed more information about the tool.

3.2.2 Gaze Added Condition

In the Gaze Added Condition participants were also able to ask questions of the robot. In addition, they were given further information about the tool they were attempting to select if they turned away from the computer display of the tools. Based on previous observations of participants, we assumed that people turned away from the task to look at the robot and ask a question. Thus, the robot elaborated on the tool when the robot sensed the participant had turned away from the task display. For example, if the participant was looking for the paring knife and turned toward the robot, the robot would respond, "The blade is smooth, not jagged."

3.2.3 Delay Added Condition

In the Delay Added Condition, participants received additional elaboration when they had asked a question or when they looked back at the robot, just as in the Gaze Added Condition. In addition to these opportunities, the robot in the Delay Added Condition provided hints when four seconds elapsed without a selection having taken place. Four seconds was the average amount of time it took a participant to select a tool in our previous experiment using the identical task. We thus assumed that when a participant hesitated for longer than four seconds, they were uncertain and could use further elaboration.

3.2.4 Immediate Added Condition

In the Immediate Added Condition, the robot offered additional elaboration when the participant asked a question or when the participant looked back at the robot, just as in the Gaze Added Condition. In the Immediate Added Condition, however, every

time the robot introduced a new tool, the robot immediately added an additional elaboration about that tool. The robot did not wait for a delay, as in the Delay Added Condition; instead, the robot included a hint in its initial turn. In this condition, for instance, the robot introduced the paring knife by saying, "Next you want a sharp paring knife. Find the paring knife. The blade is smooth, not jagged."

3.3 Procedure

When participants arrived at the experimental lab, the experimenter adjusted the gaze sensing camera to the height of each participant. Participants were informed that their gaze was being tracked. The experimenter then told the participant that the robot had been given "specific expertise" in cooking, and that "the robot will be talking to you about the tools needed to make a crème brûlée dessert." The robot was dressed to appear like a cooking expert; it wore a white chef's hat and apron and spoke with a male voice. At the beginning of the experiment, the robot was positioned in the corner of the room, about fifteen feet away from the participant. When started up, the robot opened its eyes, turned its head in the direction of the participant, and rolled over to stop about two feet from the subject. The robot explained that it will ask for the tools one at a time and reminded participants that they could ask questions if they did not know which tool to pick.

When the robot spoke, it used facial expressions. While speaking, the robot's lips moved in synchrony with its words. When speaking, the robot used basic facial expressions and turned its head to face the computer monitor whenever it referred to the tools and while waiting for a participant to select the appropriate tool. The robot spoke aloud and also displayed its messages on a display on the robot's chest. The robot used Cepstral's Theta [13] for speech synthesis, and its lips moved as it spoke. The text also showed on the screen, as in Instant Messenger interfaces. Participants interacted with the robot by typing into the same Instant Messaging interface. We used a robot without speech recognition because of current limitations in speech understanding across individuals when the dialogue is complex, as in the current case.

In the course of the dialogue, the robot prompted the participant to find cooking tools, e.g., "Find the picture of the saucepan." Pictures of several tools, the correct tool plus five incorrect tools, were shown on a nearby computer (see Figure 3). If the participants knew which tool was correct, they clicked the correct image and told the robot that they found the right tool. If the participant did not recognize the name of the cooking tool, he or she could ask the robot questions about the tool, using the IM interface. Most of participants' questions were about tool properties like shape ("does it have a round bottom"), color ("what color is it"), and usage ("what is it for"). The robot was programmed to respond to most of these inquiries. If the robot did not understand the question, it told the participant it did not know and suggested the participant ask another question. If participants made an incorrect selection, the robot informed participants they had made a mistake and asked them to try again. All the participants' tool selections and verbal responses to the robot were logged. After the participant had selected all ten tools, the robot thanked the participant and reversed its prior pattern of movement, rolling back to its original location. After conversing with the robot, the participants completed a survey about their perceptions of the robot and their conversational interaction.

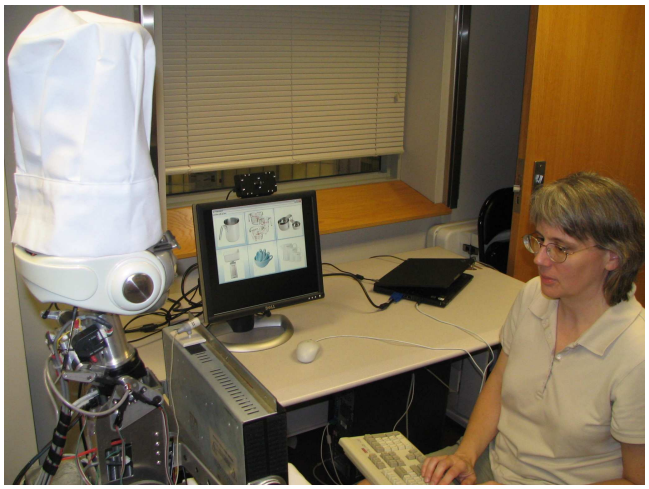

Figure 3. The experimental set-up

3.4 Dialogue Technology

The robot interpreted and responded to participants' questions using a customized variant of Artificial Intelligence Mark-up Language [16][24], a publicly available pattern-matching text processor. In previous experiments, we found that the existing implementation of AIML could not respond well to participants' questions, in part because it could not make use of dialogue context. To gain more control over the flow of the dialogue, we wrapped another technology layer around AIML and made significant changes to how AIML is processed. These modifications greatly improved the robot's ability to understand and respond intelligently. The AIML search algorithm was improved, and we created a database of properties of each tool, so when a subject asked questions like "does the saucepan have a handle", "how long is the handle", or "what color is the saucepan", the robot could answer all of these questions correctly, and many more.

For each tool, a set of elaborations was created. These elaborations were progressively offered, according to the experimental condition. The robot kept track of what questions had been asked and what elaborations had been given thus far, so that each elaboration that was offered was new information. If the participant had already asked a question about the handle of the saucepan, for example, the robot gave information that had not been discussed like the size of the saucepan or the material.

3.5 Gaze Sensing Technology

We implemented gaze sensing by mounting an eyeBox eye contact sensing camera [25][21] to the top of the monitor where participants made their selection. In Figure 2, the eyeBox is sitting on top of the monitor. The experimenter adjusted the monitor so that each participant's face was in the field of view of the eyeBox. Though it was only utilized in three of the four conditions, the eyeBox was introduced to and adjusted for participants in all conditions.

The eyeBox uses infrared light to illuminate any pupils in the frame and outputs the number of eyes it finds in the frame at a rate of ten times per second. The output also includes whether the eyes were detected looking directly toward the camera and the location of the eyes in the frame. If two eyes were detected looking directly toward the camera, the robot registered eye contact. The green

boxes in Figure 4 display direct eye contact. When direct eye contact was not available, we used the steady visibility of two eyes over two seconds before the robot would register eye contact. Sometimes eyeglasses, teeth, or hair would register as eyes, so we also required that two eyes were located at approximately the same horizontal level in the frame. These constraints greatly improved the accuracy in informal pretests with ten subjects.

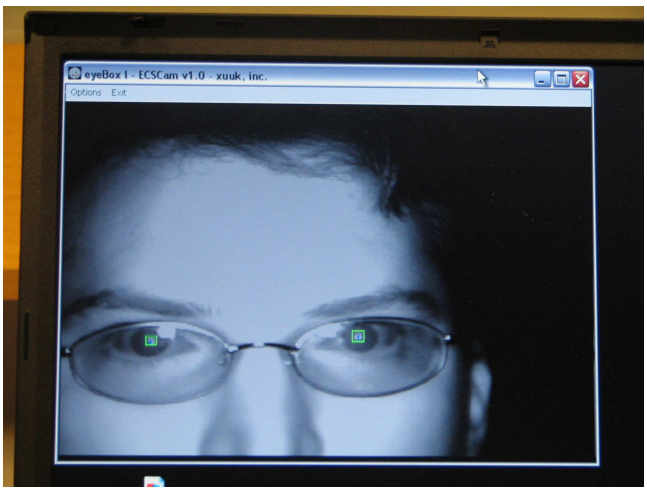

Figure 4. The visual eyeBox display

3.6 Measures

In order to be able to control for expertise in our later analyses, we measured the cooking expertise of each participant. We also collected the following outcome measures: the participants' task performance, their verbal interaction with the robot, and their subjective evaluations of the robot, the conversation, and the task.

3.6.1 Participant Expertise

To measure the cooking expertise of each participant, in the post-experiment questionnaire, we asked participants to identify eight different cooking methods. In previous work [23], we used participants' knowledge of these methods to predict their prior knowledge of the ten cooking tools in this experimental task. In a pilot survey, knowledge of the cooking tools was highly correlated with knowledge of the cooking methods, $r = .71$, $F(1,15) = 15.4$, $p = .001$. We classified participants who correctly identified five or more cooking methods as "experienced" and participants who identified fewer methods as "novices."

3.6.2 Performance

While conversing with the robot, participants were asked to select ten cooking tools. Each correct tool was presented in a group, in random position, with five incorrect tools. We counted each time a participant selected the wrong tool as a selection mistake. The total number of selection mistakes and time on task were our measures of task performance. Fewer mistakes and shorter completion times represent better performance.

3.6.3 Communication

In addition to these performance measures, we measured the number of questions participants asked while attempting to select the correct tools as well as the number of elaborations initiated by the robot. The number of questions asked describes the process by which the participant and the robot communicated. A greater number of questions indicates that more effort was required for grounding on the part of the participant.

3.6.4 Subjective Evaluation

We also used self-report questionnaire items as measures of the success of the communication process. Participants completed this questionnaire following their interaction with the robot. The questionnaire covered perceptions of the robot's characteristics (authority, intelligence), evaluations of the quality of the communication (effectiveness, responsiveness, control), and evaluations of the task (enjoyability, ease). The scales were identical to those used in [23].

Table 2. Self-report scales

Scale	Sample Questionnaire Item
Robot Authority	Expert/Inexpert
Robot Intelligence	Intelligent/Unintelligent
Robot Patronizing	My partner talks down to me.
Conversational Effectiveness	I found the conversation to be very useful and helpful.
Robot Responsiveness	My partner can adapt to changing situations.
Conversational Control	My partner dominated the conversation.
Task Enjoyability	I enjoyed participating in this task.
Task Difficulty	This task was difficult.

4. RESULTS

In this section, we explore the effects of both expertise and the condition manipulations on performance, communication, and subjective evaluation measures. Our model includes the elaboration condition, participant expertise, and the interaction between condition and expertise. In order to further determine which of the four conditions were significantly different from one another we compared all possible paired conditions using Student's t test.

4.1 Performance Measures

We first considered the effect of expertise and condition on the number of tools participants chose incorrectly before finding the correct tool (see Figure 5). There were two significant main effects and no interaction. Experts make significantly fewer mistakes than novices, $F(1,65) = 11.5$, $p = .001$ (Experts $M = 2.7$, $SD = 2.1$, Novices $M = 5.1$, $SD = 3.7$). The main effect of condition on selection mistakes is also significant, $F(3,63) = 2.98$, $p < .05$ (Question $M = 4.7$, $SD = 3.5$; Gaze $M = 5.3$, $SD = 3.5$; Delay $M = 3.5$, $SD = 3$; Immediate $M = 2.4$, $SD = 2.3$). Post hoc comparisons of all four conditions show that participants in the Immediate Added Condition do make fewer mistakes than participants in either the Questions Only or Gaze Added Conditions.

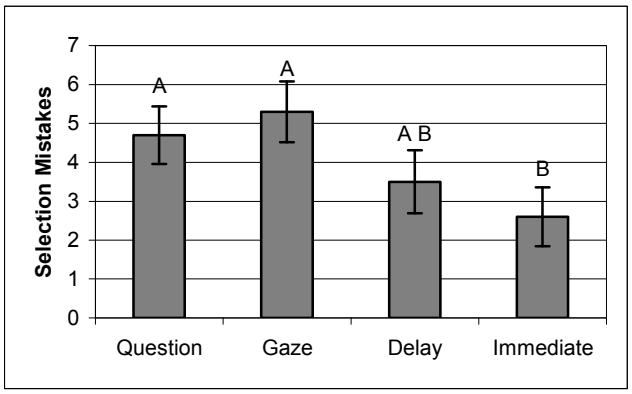

Figure 5. Number of selection mistakes by condition. (Conditions not labeled with the same letter are significantly different.)

Next we considered the effects of expertise and condition on the amount of time participants required to complete the task. There was a main effect for expertise. Experts take less time to complete the task than novices, $F(1, 65) = 15.7$, $p < .001$. An ANOVA for the effect of experimental condition on time was not significant. Participants in each condition spent roughly the same amount of time conversing with the robot. There was no significant interaction.

4.1.1 Summary

As might be expected of those more familiar with the cooking tools, experts took less time on task and made fewer mistakes than novices. There was no interaction across conditions. There were few performance differences across participants in each of the four conditions, except for the finding that participants given immediate elaborations made fewer mistakes than those who had to request elaborations, either with questions or with their gaze.

4.2 Communication Measures

Next we considered our measures of communication, including the number of questions participants asked the robot. In addition to receiving elaborations as the result of a question, participants received elaborations immediately and/or as the result of gaze or task delay (depending on their condition). Figure 6 provides an overview of how many of each type of elaboration was received in each condition. Participants in the Delay Added and Immediate Added Conditions tended to receive a greater number of elaborations overall. The robot's awareness of gaze contributed only a small number of elaborations in each condition. As shown in Figure 6, the average number of elaborations participants received as the result of task delay was more ($M = 7.1$, $SD = .6$) than the number of elaborations participants received as the result of responsiveness to gaze (gaze-prompted elaborations in the Gaze Condition $M = 1.75$, $SD = .4$; Delay Condition M = .9, $SD = .4$, Immediate Condition $M = 1.6$, $SD = .4$).

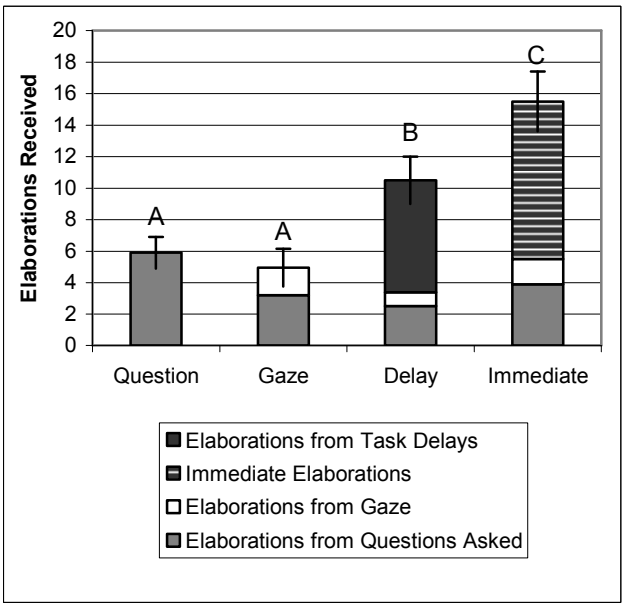

Figure 6. The average number of elaborations received in each condition. (Conditions not labeled with the same letter are significantly different.)

Our analysis of the number of questions participants ask in each condition again revealed a main effect for expertise; experts ask fewer questions of the robot than novices, $F(1,65) = 7.5$, $p < .01$ (Figure 7). There was no main effect for condition, and there was no significant interaction between condition and expertise. Student's t tests revealed that two conditions differed significantly from one another. The Questions Only Condition asked significantly more questions than the Delay Added Condition.

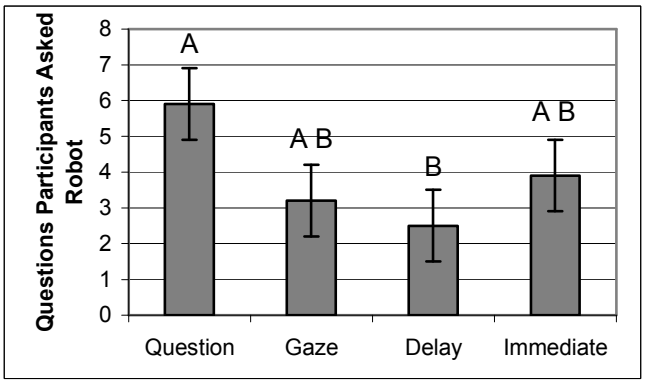

Figure 7. Number of questions participants asked by condition. (Conditions not labeled with the same letter are significantly different.)

While condition was not a significant predictor of time on task in our previous analysis, the number of questions asked is a significant factor in predicting time ($R^2 = .72$, $F(1,65) = 172$, $p < .0001$). Each question asked increased the amount of time it took participants to complete the task. There are no other significant correlations between the performance and communication measures.

4.2.1 Summary

There were no interaction effects between expertise and condition on the communication measures, but there was a significant main effect for expertise. Experts ask fewer questions than novices. The only significant difference between conditions on the communication measures is a significant decrease in the number of questions asked when the robot is responsive to task delays compared to when the robot responds only to questions.

4.3 Subjective Evaluation Measures

There were no significant differences in the subjective evaluation measures by condition. We investigated possible relationships between our performance and communication measures and the subjective evaluation measures (see Table 3). We found that the more questions a participant asked, the less control the robot was perceived to have had over the conversation, $F(1,65) = 11$, $p < .001$, and the more difficult the task was reported to be, $F(1,65) = 4.2$, $p < .05$. The more difficult participants reported the task to be, the less they enjoyed the task, $F(1,65) = 28.2$, $p < .01$.

Table 3. Correlation table

Variables	1	2	3	4	5
1. Time	1				
2. Questions Asked	.85*	1			
3. Control	-.30*	-.34*	1		
4. Task Ease	-.37*	-.26*	.10	1	
5. Task Enjoyability	-.09	-.02	.22	.56*	1

*** $p < .05$**

4.3.1 Summary

The subjective evaluations we measured here were most closely associated with participants' need to interact with the robot by asking questions. When the task was more difficult, participants had to ask a greater number of questions to complete the task and they reported lower ratings of task enjoyment.

5. DISCUSSION

In this exploratory study, we sought to investigate the impact of a robot's responsiveness to gaze and task behavior. We implemented these forms of responsiveness in simple ways and tested them in an additive way. The addition of gaze awareness alone did not seem to have much of an impact on participants' performance. One conclusion might be that gaze is not a sufficient indicator that people need further elaboration, but there are other possibilities stemming from the details of our implementation. Perhaps our implementation of gaze awareness was not sophisticated enough to respond to every turn of a participant's head, prohibiting the robot from responding appropriately. It is also possible that participants did not turn their head at all when they asked a question. Many expert typists can type without turning their head. Without detailed coding of body orientation from experiment data, we cannot draw a conclusion about the role of gaze.

In this trial, we also tested the role of task delay in providing information about when participants needed help. Added awareness of task delays did significantly reduce the number of questions that participants asked, as compared to the Questions Only condition. Providing information immediately without a delay reduced the number of mistakes as compared to the Questions Only condition. Taken together, it seems the participants did require further elaboration, but there are no consistent differences between the Delay Added and the Immediate condition upon which to draw conclusions.

5.1 Limitations

Our findings are limited by our implementation of these four types of elaboration. Although we chose four seconds as the waiting period in the Delay Condition based on the results of our earlier study, it is possible that shorter or longer delays might make this strategy of adaptation more successful. Similarly, we turned on the gaze awareness function for three of our four conditions (Immediate, Delay Added and Gaze Added, but not Questions Only). Although we did not tell participants about how the gaze sensor was being used, it is possible that the results would have been somewhat different had the Immediate and Delay Added conditions not included gaze awareness.

The impact of gaze awareness might also have been different if the participants were informed explicitly about the gaze sensing mechanism and how the robot was programmed to respond to it. There are, of course, trade-offs to training users about robotic systems, but training may be necessary as novel interaction mechanisms are introduced.

Participants in this experiment communicated with the robot by typing to the robot. Communication strategies for improving the robot's ability to infer when an individual needs help could potentially be different during speech-only communication. The use of the keyboard may alter gaze and gesture behavior, word usage, as well as the willingness of participants to initiate help-seeking conversations.

5.2 Implications for Robot Design and Research

A number of researchers in the field of human-robot interaction have recognized the importance of gaze. Many researchers have focused particularly on the idea of joint attention and the appropriate simulation of gaze activity in a robotic form [19][9][2]. Others have investigated the role of a robot's gaze as a communicative act [20][14]. Sakita et. al. use human gaze information to plan appropriate cooperative behavior for their robot [18]. Further work in this area will investigate the interpretation of human gaze as it benefits human-robot communication, and further technological developments in the field of gaze sensing will assist our understanding of the role of gaze in the grounding process between humans and robots. Although our implementation of gaze was not entirely successful, we believe further research will demonstrate its importance.

Our observations demonstrate the significance of a mixed-initiative approach to robot design. Elaborations initiated by the robot, whether immediately or after delay, were certainly useful in this selection task. If all the responsibility for obtaining help were placed on the individual, accuracy would suffer. People are not always good at knowing when they need help, so a robot which initiates some amount of elaboration is useful. Further, even if people know they need help, they may not be good at knowing how to formulate their questions and get the help they need. In these situations, a robot can use gaze behavior to gain additional awareness of the situation. In making use of the ways people

signal their state, robots can provide better help. With both conversational partners working together, the grounding process between humans and robots should become more efficient.

6. ACKNOWLEDGMENTS

This research was supported by National Science Foundation ITR project #IIS-0121426 and IGERT #DGE-0333420. The authors would also like to thank Hau-Yu Wong for help running this experiment.

7. REFERENCES

[1] Argyle, M. and Cook, M. *Gaze and Mutual Gaze*. Cambridge University Press, 1976.

[2] Brooks, R.A., Breazeal, C., Marjanovic, M., Scassellati, B., and Williamson, M.M. The cog project: Building a humanoid robot. *Lecture Notes in Computer Science* 1562 (1999), 52-87.

[3] Clark, H. *Using Language*. Cambridge University Press, 1996.

[4] Clark, H. and Krych, M. Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50 (2004), 62-81.

[5] Clark, H. and Wilkes-Gibbs, D. Referring as a collaborative process. *Cognition*, 22 (1986), 1-39.

[6] Duncan, S. and Fiske, D.W. *Face-to-face interaction: Research, methods and theory*. Erlbaum, Hilldale, NJ, 1977.

[7] Fussell, S. and Krauss, R. Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, 62 (1992), 378-391.

[8] Gergle, D., Kraut, R. E., and Fussell, S. R. Action as language in a shared visual space. *Proceedings of CSCW 2004*, 2004, 487-496.

[9] Imai, M., Kanda, T., Ono, T., Ishiguro, H. and Mase, K. Robot-mediated round table: Analysis of the effect of robots gaze. *Proceedings of 11th IEEE International Workshop on Robot and Human Communication* (ROMAN 2002), 411-416.

[10] Isaacs, E. and Clark, H. References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116 (1987), 26-37.

[11] Kleinke, C.L. Gaze and eye contact: A research review, *Psychological Bulletin* (1986), 78-100.

[12] Kraut, R., Fussell, S. and Siegel, J. Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction*, 18 (2003), 13-49.

[13] Lenzo, K.A., and Black, A.W., Theta, Cepstral, http://www.cepstral.com.

[14] Mutlu, B., Hodgins, J., and Forlizzi, J. A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior. *Under review*.

[15] Nakano, Y., Reinstein, G., Stocky, T., and Cassell, J. Toward a Model of Face-to-Face Grounding. *Proceedings of ACL 2003*, 2003.

[16] Powers, A. An easy to use dialogue tool: AIMLE. Unpublished manuscript. Obtained from the author at HCII, Carnegie Mellon University, Pittsburgh, PA.

[17] Sacks, H., Schegloff, E. and Jefferson, G. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 4 (1974) 696-735.

[18] Sakita, K., Ogawara, K., Murakami, S., Kawamura, K., Ikeuchi, K. Flexible Cooperation between Human and Robot by interpreting Human intention from Gaze Information. *Proc. IEEE/RSJ Int. Conf. on Intelligent Robot and Systems (IROS)*, 2004, 846-851.

[19] Scassellati, B. Mechanisms of shared attention for a humanoid robot. *Embodied Cognition and Action: Papers from the 1996 AAAI Fall Symposium*. AAAI Press, 1996.

[20] Sidner, C., Lee, C., Kidd, C. and Lesh, N. Explorations in engagement for humans and robots. *Proceedings of the International Conference on Humanoid Robots 2004*.

[21] Smith, J.D., Vertegaal, R., and Sohn, C. ViewPointer: Lightweight calibration-free eye tracking for ubiquitous handsfree deixis. *Proceedings of UIST 2005*. (Seattle, WA) 2005, 53-61.

[22] Schober, M. and Brennan, S. Processes of interactive spoken discourse: The role of the partner. In Graesser, A., Gernsbacher, M. and Goldman, S. eds. *The Handbook of Discourse Processes*, Lawrence Erlbaum, Mahwah, NJ, 2003, 123-164

[23] Torrey, C. Powers, A., Marge, M., Fussell, S., and Kiesler, S. Effects of adaptive robot dialogue on information exchange and social relation. *Proceedings of the Conference on Human-Robot Interaction 2006*. (Salt Lake City, March 1-3), 2006, 126-133.

[24] Wallace, R. A.L.I.C.E. *ALICE Artificial Intelligence Foundation*. http://www.alicebot.org.

[25] Xuuk, inc., eyeBox. http://www.xuuk.com.