Robot-Directed Speech:

Using Language to Assess First-Time Users' Conceptualizations of a Robot

Sarah Kriz
Department of Human Centered Design & Engineering
University of Washington
Seattle, WA, USA
kriz@u.washington.edu

Gregory Anderson
Department of Psychology, ARCH Lab
George Mason University
Fairfax, VA, USA
ganders5@gmu.edu

J. Gregory Trafton
Navy Center for Applied Research in Artificial Intelligence
U.S. Naval Research Laboratory
Washington, DC, USA
greg.trafton@nrl.navy.mil

***Abstract*— It is expected that in the near-future people will have daily natural language interactions with robots. However, we know very little about how users feel they should talk to robots, especially users who have never before interacted with a robot. The present study evaluated first-time users' expectations about a robot's cognitive and communicative capabilities by comparing robot-directed speech to the way in which participants talked to a human partner. The results indicate that participants spoke more loudly, raised their pitch, and hyperarticulated their messages when they spoke to the robot, suggesting that they viewed the robot as having low linguistic competence. However, utterances show that speakers often assumed that the robot had humanlike cognitive capabilities. The results suggest that while first-time users were concerned with the fragility of the robot's speech recognition system, they believed that the robot had extremely strong information processing capabilities.**

Keywords- human-robot communication; spatial language; natural language; experimentation

I. INTRODUCTION

One vision of the future is that service robots will help humans with daily activities. Grocery shopping, picking up dry cleaning, and checking in at the doctor's office will one day involve interactions with robots. However, most people have never interacted with a robot and have little understanding of the current state of artificial intelligence. In fact, most people's experience with robots is limited to the fictional robots they have seen in film and on television. This lack of real-world experience with robots may be problematic, especially if naïve users overestimate the capabilities of a robot. Disillusionment, confusion, and frustration may prevent first-time users from adopting robotic technologies. Specifically, expectations about communicative capabilities can greatly impact the success of verbal and non-verbal interactions between a human and a robot, especially if the expectations do not match with the robot's competencies.

The objective of this study is to determine how first-time users conceptualize the cognitive and communicative capabilities of a humanoid robot. This study examines the speech modifications that people make when they talk to a robot, and looks at what those adjustments can tell us about people's conceptualizations of robots. Using what is known about human-directed speech, we assess the features of robot-directed speech to explore the expectations that first-time users have about a robot's communicative capabilities.

II. USING SPEECH MODIFICATIONS TO ILLUMINATE CONCEPTUALIZATIONS

People's conceptualizations of robots seem to be based on a number of factors, including the robot's external features [1], where they believe the robot was manufactured [2], and the amount of interaction people have with the system [3]. Previous research on conceptualizations of robots has generally been conducted using video presentations of someone interacting with a robot. Although these experiments are designed to study users' reactions or assessments of a robot's behavior, personality, or motivations, most do not allow participants the opportunity to interact with a real robot. While watching a video may provide some information on which to build conceptualizations, those conceptualizations may be different from those developed during an interaction with a robot. If interactions with robots will be common in the future, it is desirable to collect data that can evaluate people's conceptualizations about robots during face-to-face interactions.

A second methodological problem common to many previous studies on conceptualizations of robots is reliance on questionnaires to assess mental models. A general problem with using questionnaires is that they capture post hoc reflections. In terms of previous research on conceptualizations of robots, questionnaires require respondents to report thoughts and feelings they had while interacting with the robot or viewing a video. Participants must judge their emotions, strategies and perceptions retrospectively, and the act of remembering and responding to questions about previous events puts extra cognitive burden on respondents.

978-1-4244-4893-7/10/$25.00 © 2010 IEEE

Furthermore, reflections on behavior may cause people to over-evaluate their actions or feelings as a means of explaining behavior for which, at the time, they had no conscious motives.

There have been recent attempts to develop novel ways to evaluate conceptualizations and measure reactions to robot behavior in real time. Koay et al. [4] have attempted to avoid the post hoc reflection problem by developing a "comfort level device" that participants can use during their interactions. The device allows participants to make real-time judgments of their level of comfort during an interaction with the robot. While Koay et al. [4] have been able to collect data beyond what could be ascertained from questionnaire data alone, they point out several shortcomings of the comfort level device as a methodology. One is the fact that participants must make meta-cognitive decisions during the interaction. This is problematic because people's judgments could be biased on what they think they are supposed to do.

Participants may not consciously know why they acted in a certain manner or felt a certain way, yet in human-robot interaction studies they are asked to explain their thoughts, emotions, and behavior. This is especially problematic for studying communicative expectations, as so much of human communication is unconscious and automatic. In this paper we propose a methodology for studying conceptualizations of robots that does not involve post hoc reflections or meta-cognitive judgments. Drawing from research on language acquisition, we have adopted a methodology used in linguistics and psychology that evaluates users' conceptualizations of robots based on the way in which they produce language *during* interactions with a robot.

Decades of research in psychology and linguistics have examined aspects of people's mental representations through evaluation of speech and discourse (e.g., [5, 6]). A large body of research has shown that speakers adapt the way in which they speak to meet the needs of listeners, and that speech modifications can illuminate speakers' conceptualizations of their listeners' cognitive abilities and perceived communicative needs [7-9].

Much of what is known about adjustments people make in their speech and language comes from the study of infant-directed speech. When speaking to infants, people tend to change a number of characteristics of their speech, such as pitch, loudness and clarity, in order to promote successful communication. These systematic modifications constitute a speech "register" that differs from standard speech to adults according to the perceived needs of the infant. For example, the infant-directed speech register is generally characterized by elevated speaking pitch and exaggerated pitch contours [10-12]. It is theorized that the "singsong" quality resulting from elevated pitch and high pitch variability serves the purpose of gaining the attention and communicating emotion. Therefore, the way that adults modify their speech provides insight into their perception of infants' cognitive, social, and linguistic needs; in the case of infant-directed speech, adults perceive the infant to have limited attention and a preference to attend to emotionality in speech.

A second characteristic of infant-directed speech is hyperarticulation [8], or careful pronunciation of certain sounds in words. For instance, in normal conversational speech the [t] in the word 'can't' is often produced without aspiration (the high frequency burst of air at the end of the sound), making "can't" sound similar to its opposite, 'can.' Normally, the lack of aspiration does not have a negative effect on a competent speaker of English because she will use context and unconscious linguistic knowledge to identify the word. However, less competent speakers of English may have trouble identifying the unaspirated [t]. In infant-directed speech, speakers tend to hyperarticulate, making sure to produce the aspiration when producing the [t] sound. This hyperarticulation clarifies the individual sounds in a word. It has also been hypothesized that hyperarticulation serves a didactic purpose [8]. That is, by hyperarticulating, adults are attempting to help the infant learn the words and sounds of the language.

In addition to changing the way they produce sounds when speaking to infants, adults often also lower the informational complexity of a message. Obviously, infants have limited vocabulary and memory resources. To compensate, adults use primarily single words or short phrases with repetitions of high content words, rather than complex, connected sentences.

Table I summarizes what, in American culture, are perceived to be the communication needs of infants, and the modifications adults make to their speech in order to accommodate those needs.

TABLE I. INFANTS' PERCEIVED COMMUNICATION NEEDS AND ASSOCIATED SPEECH MODIFICATIONS

Perceived Need	Speech Modifications
Attentional	Increased loudness, elevated pitch
Affective	Increased pitch variability
Linguistic/Didactic	Hyperarticulation
Cognitive	Simplified grammar and vocabulary, shorter utterances

Speech adjustments such as the ones described above are made not only for infants. When speaking to pets, adults exhibit the same pitch elevations and fluctuations that they do with infants. There is no evidence, however, of hyperarticulation during speech directed at pets [13]. This is hypothesized to be because most pets are not expected to be able to learn the sounds of language, and therefore this didactic modification, which is present in infant-directed speech, is not a feature of pet-directed speech.

Conversely, when talking to non-native adult conversation partners, speakers exhibit the hyperarticulation characteristic of infant-directed speech, but not the elevated pitch or exaggerated pitch contours [9]. This is consistent with the hypothesized purpose of pitch and pitch variability as attention-getting devices. Adult non-native speakers do not generally have the attentional and affective needs of infants. They are, however, trying to learn the sounds of the new language, thus speakers seem to accommodate that need through hyperarticulation.

Simplified grammatical structures and vocabulary are also not unique to infant-directed speech. These adjustments have been demonstrated in speech directed to children, non-native listeners, and people identified as having cognitive impairments [14, 15].

Taken as a whole, this body of research suggests that people speak "normally" to others who are perceived to have adult-like cognitive and linguistic abilities. However, when there is a perceived potential difficulty, speakers systematically change their speech patterns to accommodate the needs of their communication partner. These changes in speech patterns provide information regarding the expectations speakers have about their addressee's cognitive and linguistic capabilities.

While speech modifications have been used to study how speakers adapt to human populations with differing cognitive, attentional, and linguistic needs, there is reason to believe that they may also illuminate conceptualizations of the communicative capabilities of non-human populations, including robots. In addition to communicating with humans and pets, most people have experience talking with a range of newer technologies, from voice recognition dialing commands on cell phones to telephone menu systems. Each of these items is different from the others in its ability to process verbal input, and it is likely that the often unconscious language adaptations made when communicating with infants, animals, and non-native speakers occur when talking to non-human addressees as well. We propose that studying speech modifications in human-robot communication can provide insight into how people, especially first-time users, conceptualize the communicative needs of a robot.

III. Empirical Study

Robots are quite different from human populations, and it is uncertain how people might conceptualize the attentional, emotional, and linguistic needs of robots. People could speak to robots in several different ways. They could talk to a robot the same way that people talk to their colleagues. Alternatively, they could talk to robots the same way they would talk to children, pets, or non-native speakers. A final possibility is that robots could be in a class different from all the above.

Based on previous work in infant-, child-, and non-native-directed speech, a number of predictions can be made regarding how people with no prior experience with robots might speak to a robot. Assuming that speakers view robots as a population that requires extra clarity and attentional cuing, their speech should reflect those beliefs. Specifically, speakers should increase in pitch and loudness when they direct their speech to a robot. They may also hyperarticulate sounds to provide the robot with clearer speech input. Furthermore, if speakers expect a robot to have limited language comprehension capabilities, they should simplify their language by using fewer words to describe a target object, simpler grammatical structures and vocabulary, and a slower rate of speech when speaking to the robot.

These predictions were tested in an experimental setting in which participants spoke to a robot. Their speech was compared to a second session, in which they spoke to a human while completing the same task.

A. Method

1) Participants

Fifteen American students (8 female, 7 male) from the University of Washington participated in the experiment. None of the participants had previously talked to or interacted with a robot, and all 15 were native speakers of English.

2) Procedure

The experiment employed a within-subjects design, in which participants visited the laboratory twice and completed a referential communication task with a robot and with a human. The order of the sessions was counterbalanced across participants so that half of the subjects had their first session with the robot and the other half completed the human condition first.

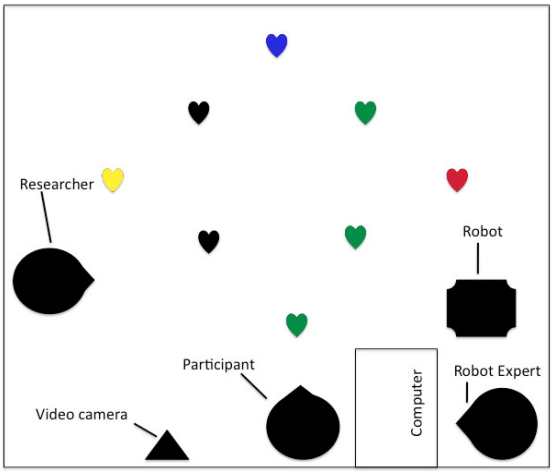

Figure 1. Experimental setup.

In both sessions participants were greeted in the laboratory by a researcher and were introduced to a confederate who acted as the "robot expert." The researcher explained that the robot expert knew how to program the robot and was there to help out during the experiment. An array of eight shapes of various colors was arranged on the floor roughly two feet from where the participants sat. (See Figures 1 and 2.) Participants were handed a booklet that contained pictures matching the array displayed on the floor. In each picture, one target object was circled. In the robot condition, participants were asked to direct the robot to the circled object. In the human condition, participants gave their commands to the robot expert, who guided the robot to the circled object.

In both the human and robot sessions, a third researcher, the 'wizard,' was in an adjacent room watching and listening to participants through a hidden video feed. During each trial, the wizard guided the robot to the correct item, regardless of the requests participants gave. The robot expert confederate in the testing room did not have any control over the robot. Thus, the two sessions were exactly the same except that participants *believed* they were talking to a robot in one case and to a human in the other.

Figure 2. Picture of array of objects and robot used in the experiment.

During the robot session, participants were told that the robot expert would put the robot into "autonomous mode." (Although the wizard in the control room actually changed the screen text, the robot expert typed something on the computer in the testing room to make it look like she was controlling what was displayed on the screen.) The robot's screen changed from showing the phrase, "Waiting for user input" to displaying the words "Autonomous Mode." The robot expert then stepped away from her computer and stood in a corner of the room for the rest of the session. The participants instructed the robot to navigate to the circled object in each picture, and they talked into a head-mounted microphone that they were told fed into the robot's computer.

In the human session, participants were instructed they needed to tell the robot expert where to direct the robot, according to the circled object in each picture. Before they began the task, the robot expert pretended to change the robot's screen from "Waiting for user input" to displaying the words "Researcher Control Mode." During the trials, participants talked into the head-mounted microphone, and the robot expert wore earphones to listen to the participants' requests. The robot expert typed participants' commands into a text editor to make it look like she was controlling the robot to the specified object.

Sessions were scheduled 1-7 days apart, according to individual participants' personal obligations. All sessions were video and audio taped. At the end of the second session, participants filled out a written questionnaire that asked about their age, gender, and the extent to which they use related technologies such as the internet and cell phones.

3) Data Analysis

Digital audio recordings were segmented into individual trials for analysis. Requests were defined as the complete command given for each trial, including immediate corrections ("Get me the green circle next to, I mean to the right of, the blue circle"), and filled pauses ("Get me the, *um*, green…) while excluding filled pauses prior to the first word of the command ("*Um*, get me the green…).

The recordings for each condition for each subject were normalized as a group to maximize acoustic energy available for pitch and duration analyses, while preserving relative loudness levels between conditions. After normalization, recordings were segmented into individual trials. Trials were then individually analyzed for pitch, pitch variability, loudness, and rate of speech using the Praat acoustic analysis software [16]. Intensity (loudness) and frequency (pitch) measures were computed using Praat, with the pitch floor and ceiling set differently for men and women to minimize artifacts. Copies of the wav files were filtered with a low-pass filter set at 400Hz prior to pitch analyses. Loudness was measured in decibels, and pitch was measured using semitones calculated from an initial value of 1 Hz, allowing the use of parametric statistical procedures for data analysis. To determine rate of speech, the number of syllables per second was calculated for each request by measuring the duration of the requests using the duration function in Praat. The number of syllables was then divided by the duration of the request in seconds.

To determine the complexity of the informational content of utterances, only the object noun phrases were analyzed. In other words, for a request such as 'Go to the green heart on the right' only the words contained in the noun phrase 'the green heart on the right' were counted. Word counts were based on common orthographic conventions in American English.

To measure hyperarticulation, speakers' requests were coded to determine the frequency with which they aspirated word-final [t]. Aspiration is a puff of air that is produced after a consonant, in this case, [t]. In causal conversational American English, word-initial [t] is aspirated (e.g., tiny), but word-final [t] is not. However, when speakers want to be sure that that a word-final [t] is perceivable, they will hyperarticulate the sound by aspirating the [t].

A coder with a Masters degree in psycholinguistics listened to all the trials for all 15 participants. The coder received the audio files in a random order and was blind to speaker, condition, experiment methodology, and hypotheses. She noted all cases in which speakers aspirated word-final [t]. The number of aspirated cases for each participant's robot and human sessions were divided by the total number of word-final [t] environments found in each session, yielding the percentage of aspirated word-final [t] for each session. As a reliability check, the first author also coded aspiration for one-third of the trials. Of the 149 cases of word-final [t], the two coders agreed on 94% of the cases.

B. Results

A mixed-model ANOVA was conducted for several analyses below. A mixed-model ANOVA is used when there is at least one between variable (e.g., condition) and at least one within variable (e.g., order, gender). Participant was considered a random variable. Mixed-model ANOVAs were conducted for each independent variable to determine main effects of condition, gender, and order. However, no gender or order effects were found. For ease of reading, statistics from paired-samples t-tests (robot v. human condition) are reported below.

1) Pitch

Pitch values significantly differed across condition [t (14) = 3.05, p = .01]. Speakers used significantly higher pitch (M =

87.69, *SE* = 1.33) when talking to the robot than when talking to the human researcher (*M* = 86.97, *SE* = 1.42). Figure 3 shows the individual participants' data in terms of the extent to which they raised their pitch when talking to the robot. The bars extending above the x-axis show the increase in pitch (in semitones) from the human to robot conditions. Bars below the x-axis represent participants who exhibited higher pitch when talking to the human than the robot. The majority of participants raised their pitch when talking to the robot.

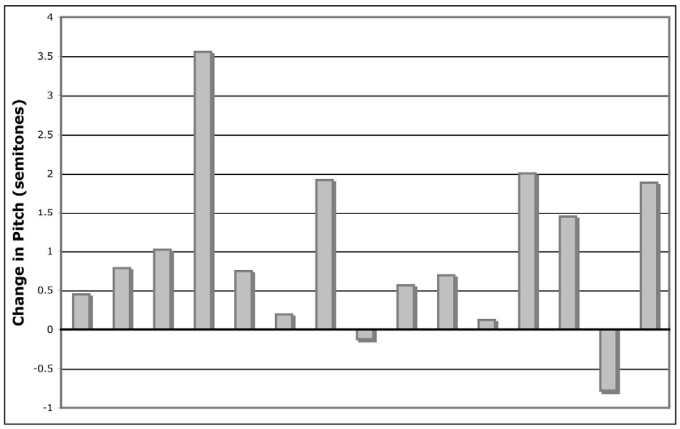

Figure 3. Individual participants' mean change of pitch from human to robot conditions. Positive change scores represent an increase in pitch when talking to the robot.

2) Pitch Variability

Speakers did not differ the extent to which they varied their pitch when talking to the robot (*M* = 3.73, *SE* = .28) versus talking to the researcher (*M* = 4.17, *SE* = .38; [*t* (14) = -1.21, n.s.].

3) Loudness

Participants were louder when addressing the robot than when talking to the researcher [*t* (14) = 3.45, *p* < .01]. Although participants spoke into the head mounted microphone in both conditions, they tended to speak louder when they thought that the robot was processing their speech (*M* = 72.32, *SE* = .95) than when they were talking to the researcher (*M* = 68.51, *SE* = 1.38). Figure 4 shows individuals' change data when comparing the human condition to the robot condition. As the figure shows, 13 of the 15 participants exhibited louder speech patterns when they talked to the robot, some raising their amplitude on the order of 10 dB or more.

4) Rate of Speech

Rate of speech was predicted to be slower when speaking to the robot. However, contrary to the prediction, subjects' rate of speech when talking to the robot was not significantly slower than when talking to the human researcher (Robot: *M* = 4.22, *SE* = .23 v. Human: *M* = 4.26, *SE* = .24; [*t* (14) = -0.22, n.s.]).

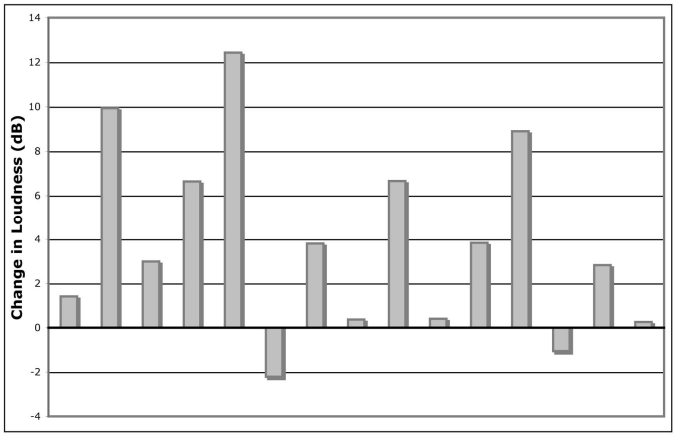

Figure 4. Individual participants' mean change of loudness from human to robot conditions. Positive change scores represent an increase in loudness when talking to the robot.

5) Hyperarticulation

As discussed previously, aspirating word-final [t] is sign of hyperarticulation. As predicted, participants hyperarticulated more often when talking to the robot than when talking to the researcher. Speakers aspirated 25.91% of word-final [t] (*SE* = 7.49%) when talking to the robot, compared to 13.09% (*SE* = 3.70%) when talking to the human [*t* (14) = 2.13, *p* = .05]. These results seem to reflect the speakers' desire to produce a clear message for the robot.

6) Conceptual Complexity of Requests

When describing the target object, subjects used more words when speaking to the robot (*M* = 6.65, *SE* = .36) than when describing the object to the human (*M* = 6.12, *SE* = .36; [*t* (14) = 3.50, *p* < .01]). The quantitative differences between the informational content across the two conditions was manifested in different ways. Some speakers provided more spatial information to the robot than to the human. This ranged from adding relatively unhelpful spatial details to robot-directed requests (i.e., specifying the target object was 'on the floor' when talking to the robot) to adding complex spatial information such as the target object's relation to a landmark that was not mentioned in the equivalent researcher-directed trial. Another way in which some speakers changed the complexity of their requests to accommodate the robot was through the use of full syntactic structures. For instance, in the human-directed condition, one participant referred to the target object using a reduced relative clause, "the black heart on the left side," omitting the complementizer 'that.' When talking to the robot, the speaker referred to the same target object as "the black heart *that's* on the left side of the yellow heart." In addition to specifying a reference object when talking to the robot, the speaker also added the complementizer 'that' to create a full relative clause. This example illustrates how additional words sometimes served to provide more spatial information, but were also used to create fuller, clearer grammatical structures.

C. Implicit Expectations about Robots as Communicative Partners

Requests to the robot were also qualitatively analyzed to determine implicit expectations about the robot's cognitive capabilities. Based on the information speakers commonly gave in their robot-directed requests, several basic expectations seem to be consistent across speakers. Mostly notably, speakers assumed that the robot could determine quite complex spatial relations. Table II provides a list of the most common spatial relations found in the robot-directed trials. (Note that all the terms listed in the table were used by at least one-third of the participants.) The most frequent spatial relations used when speaking to the robot were 'closest/nearest/furthest.' Eleven of the 15 participants used these terms and they appeared in 28% of the requests to the robot. Other spatial terms that were used frequently were 'between,' 'left/right,' and 'next to.'

Table II also shows the spatial relations used most frequently when talking to the human researcher. A comparison between the two conditions indicates that speakers used the same types of spatial terms in roughly the same proportion when talking to the robot and the human researcher, suggesting that speakers felt the robot could compute complex spatial relations that are normally used when talking to cognitively competent adults.

TABLE II. MOST FREQUENT SPATIAL TERMS CONTAINED IN ROBOT-DIRECTED AND HUMAN-DIRECTED TRIALS

Spatial term	Percent frequency	
	Robot-directed trials	Human-directed trials
Closest/nearest/furthest	27.5%	25.8%
Between	14.2%	12.5%
Right/left	12.5%	8.3%
Next to	6.7%	10%

Computing fuzzy spatial relations such as 'next to,' 'near' or 'to the left,' although easy for humans, is a difficult task for robots (e.g., [17-19]). Yet, many speakers assumed that the robot could calculate an object's location based on these fuzzy relations. Furthermore, requests containing many of these terms provided little information on the perspective from which to calculate them. For example, speakers very commonly told the robot that an object was 'to the left,' but did not specify a reference object from which to calculate the term 'left.' Many possibilities exist: the left of the speaker, the left of the robot, the leftmost area of the array.

In some cases participants did seem concerned with providing the robot with additional referential information, and this too can be taken as evidence of expectations of high cognitive functioning. Speakers often assumed that the robot was flexible in calculating spatial relations from a variety of reference objects. Across the eight trials, individuals often switched between using themselves, the robot, and other objects in the room as reference objects. For instance, throughout the robot-directed trials, one speaker switched reference objects three times. Over the eight trials, he went from using another object in the array ('closest to *the blue heart*'), to the researcher ('closest to *Jonathan*'), and finally, to the speaker himself ('closest to *me*'). This use of multiple reference objects suggests that speakers believed the robot had the flexibility to identify and keep track of several objects and update their positions in relation to its own and other people's current spatial locations.

While the requests directed to the robot suggest that people had high expectations about the robot's ability to calculate spatial relations, there is evidence that they had equally high expectations regarding the communicative capabilities of the system. Here we feel it is important to divide the domain of language into low-level speech and higher-level communicative abilities. As discussed earlier, it seems that participants were wary of the robot's speech recognition capabilities, as they exhibited increases in hyperarticulation, pitch, and loudness. However, the data suggest that many speakers assume the robot was capable of doing high-level communicative tasks such as differentiating when speech was directed to it versus someone else in the room, and remembering what had been said previously. For instance, many of the participants made comments to the researchers in the room between trials while wearing the head-mounted microphone, presumably thinking that the robot would not process the speech they were directing to other people.

Similarly, two participants seemed to assume that the robot had short-term memory and could remember what they had said on previous trials. One participant started every request for trials 2-8 with the words 'now' or 'next,' as if the robot was aware that it was doing a sequence of trials and had completed a trial previously. Further evidence comes from another participant, who described the target object on one trial as 'the other black heart.' The participant had already asked the robot to navigate to one of the two black hearts, and when the other black heart appeared as the target object she opted not to provide a spatial description and instead referred to it as the 'other' black heart, suggesting that she expected the robot to remember the black heart in the previous trial.

D. Discussion

The results of this study show that first-time users do not talk to a robot like they speak to a fellow adult human. The quantitative speech measures suggest that participants attempted to make their speech clearer and easier to understand. They raised their pitch and loudness when they believed they were talking to the robot. Furthermore, they hyperarticulated word-final [t] more often when talking to the robot. These modifications were likely done to ensure the robot's speech recognition system could recognize the sounds and words they produced. The significant increase in the number of words speakers used to describe the target object also suggests that they were concerned about the fragility of the robot's language processing system. Several speakers added additional spatial information when talking to the robot, and a few participants changed the syntactic structure of their utterances to include complementizers and words that are commonly omitted when talking to adult English speakers.

One of the interesting patterns that emerged from these results is the apparent contradiction between the production of careful speech and the number of words used to describe the target object. At first these results might indeed seem to be a

contradiction—if a person is concerned about the reliability of a speech recognition system, using more words creates a higher probability for failure. However, if the speaker feels that the extra words aid in informational clarity, then they are worth the added cost of a speech recognition error. Thus, the results suggest a trade-off between speech recognition and information processing. Based on the human behavior measured in the experiment, it seems that speakers strive to produce phonetically and informationally clear language when talking to a robot.

From an engineering perspective, the findings of this study can be helpful to designers and engineers who work on the design of speech recognition and communication systems for robots. Normally, developers of such systems use human-to-human speech as a baseline. However, as our findings indicate, speakers do not talk to robots as they talk to other humans. Therefore, the "normal" human range of speech may be a misleading starting point when developing of speech recognition systems for robots. The naturalness and success of an interaction with a robotic system depends on whether the system's speech and language capabilities match the way in which people think they should talk to the robot, and therefore understanding how people talk to robots is crucial to the engineering process.

Additionally, the results of this study remind us that communication is not simply about speech recognition and language parsing. Users are constantly juggling goals of providing adequate information to multiple input systems. The patterns that emerge from the data suggest that while trying to produce clear input for the speech recognition system, speakers are also concerned with providing adequate spatial and referential information so that the robot will be able to identify the correct referent. This trade-off shows how important it is to acknowledge the complex relations that exist between a robot's various computational systems and to consider how a user's complex communicative expectations will cause different input systems to interact with each other.

The qualitative analysis of the content of speakers' requests to the robot provides further support of complex communicative expectations. The participants' utterances clearly show that although they did not expect the robot to have superior speech recognition, they did expect it to have cognitive capacities that are on par with human cognition. While we cannot be absolutely certain about the source of these first-time users' implicit expectations, some of these expectations can be linked to the task itself. First, it seems reasonable that participants assumed that the robot could understand English and differentiate between colors based on the design of the task (i.e., instruct the robot to navigate to the target shape). Additionally, that the robot always responded correctly probably encouraged participants to build unrealistic expectations about the robot's higher-level language capabilities. However, expectations of short term memory and the ability to determine the intended audience of a message are most likely due to an unconscious understanding of the fundamental elements of human communication and social cognition. This understanding may be extended to robots by default. Because the current study was not designed to evaluate this claim with any level of certainty, more research must be conducted to identify whether and when people extend beliefs about human communicative abilities to robot partners.

Another area for future research concerns how the physical form of a robot may influence speakers. In a previous pilot experiment [20] we asked participants to complete a similar task with a Sony AIBO robot, which has a dog-like appearance. Many of the results were quite different. Notably, several speakers used telegraphic utterances such as 'Fetch blue heart' when speaking to Aibo. It is possible that the shape, movements, and barking behavior of Aibo may have evoked expectations that are more in line with dogs than with more humanoid looking robots. Because the two experiments collected data from different populations, we must take care in comparing across the experiments.

The results of this study as well as our previous work suggest that the design of a robot may strongly affect people's conceptualizations about its cognitive and communicative capabilities. Although at this point it is premature to specify definitive design guidelines, this line of research suggests that the physical form of a robot (as well as a robot's behaviors) may evoke certain expectations about the robot's cognitive and communicative capabilities.

IV. CONCLUSIONS

This study examined the features of speech directed to a robot and how these features are similar and different to speech directed to a human, given the same task and same constraints. First-time users do not talk to robots as they talk to human adults, although they do seem to expect that robots will have the same high-level communication competencies that humans exhibit. In this paper we have shown that the evaluation of robot-directed language is a powerful methodological tool that can provide insights into people's implicit beliefs about a robot's cognitive and linguistic capabilities.

ACKNOWLEDGMENT

The authors would like to thank Jonathan Morgan, Priya Guruprakash Rao, and Toni Ferro for their help in data collection, as well as Ian Finder and Eli White for their programming expertise. Gratitude is also due to Shannon Riddle for her hyperarticulation coding. This work was partially supported by the Office of Naval Research under funding documents N0001409WX20173 and N0001409WX20412 to JGT. The views and conclusions contained in this document should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Navy.

REFERENCES

[1] Kiesler, S. & Goetz, J. 2002. Mental models of robotic assistants. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (Minneapolis, MN, USA, April 20-25, 2002) CHI '02. ACM, New York, NY, 576-577.

[2] Lee, S-L., Kiesler, S., Lau, I. Y-M., & Chiu, C-Y. 2005. Human mental models of humanoid robots. In *Proceedings of the International Conference on Robotics and Automation* (Barcelona,

April 18-22, 2005). ICRA 2005. IEEE, Piscataway, NJ, 2767-2772.

[3] Fussell, S., Kiesler, S., Setlock, L., & Yew, V. 2008. How people anthropomophize robots. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam, The Netherlands, March 12-15, 2008). HRI '08. ACM, New York, NY, 145-152.

[4] Koay, K., Dautenhahn, K., Woods, S., & Walters, M. 2006. Empirical results from using a comfort level device in human-robot interaction studies. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction* (Salt Lake City, Utah, USA, March 2-3, 2006). HRI '06. ACM, New York, NY, 194-201.

[5] Chafe, W. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production.* Norwood, NJ: Ablex.

[6] Levinson, S. 2003. *Spatial language and cognition.* Cambridge: Cambridge University Press.

[7] Freed, B. 1980. Talking to foreigners versus talking to children: Similarities and differences. Research in second language acquisition: Selected papers of the Los Angeles Second Language Acquisition Research Forum.

[8] Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. 1997. Cross-Language Analysis of Phonetic Units in Language Addressed to Infants. *Science, 277*(5326),684.

[9] Uther, M., Knoll, M. A., Burnham, D. 2007. Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication, 49*, 2-7.

[10] Garnica, O. K. 1977. Some prosodic and paralinguistic features of speech to young children. *Talking to children: Language input and acquisition*, 63–88.

[11] Fernald, A., & Simon, T. 1984. Expanded intonation contours in mothers' speech to newborns. *Developmental psychology, 20*(1), 104-113.

[12] Fernald, A. 1989. Intonation and communicative intent in mother's speech to infants: is the melody the message? *Child development, 60*(6), 1497-1510.

[13] Burnham, D., Kitamura, C., & Vollmer-Conna, U. 2002. What's New, Pussycat? On Talking to Babies and Animals. *Science, 296*(5572), 1435-1435.

[14] Snow, C. E. 1972. Mothers' speech to children learning language. *Child Development, 43*(2), 549-565.

[15] Depaulo, B. M., & Coleman, L. M. 1986. Talking to children, foreigners, and retarded adults. *Journal of personality and social psychology, 51*(5), 945-959.

[16] Boersma, P. & Weenink, D. 2008. Praat; doing phonetics by computer (Version 5.0.06).

[17] Blisard, S. N. & Skubic, M. 2005. Modelling spatial referencing language for human-robot interactions. *Proceedings of IEEE International Proceedings on Robot and Human Interactive Communication 2005. ROMAN 2005*, 698-703.

[18] Skubic, M., Perzanowski, D., Blisard, S., Schultz, A., Adams, W., Bugajska, M. & Brock, D. 2004. Spatial language for human-robot dialogs. *IEEE Transactions on Systems, Man, & Cybernetics, 34*(2), 154-167.

[19] Tenbrink, T. 2003. Communicative Aspects of Human-Robot Interaction. In: Helle Metslang & Mart Rannut (Eds.), *Languages in development*, Lincom Europa.

[20] Kriz, S., Anderson, G., Bugajska, M., & Trafton, J. G. (2009). Robot-directed speech as a means of exploring conceptualizations of robots. In *Proceedings of the ACM/IEEE International Conference on Human Robot Interaction* (San Diego, CA, March 11-13, 2009), HRI '09. ACM, New York, NY, 271-272.