Human Social Response Toward Humanoid Robot's Head and Facial Features

Jun Ki Lee
MIT Media Laboratory
20 Ames St., E15-468D
Cambridge, MA 02139 USA
joon@media.mit.edu

Cynthia Breazeal
MIT Media Laboratory
20 Ames St., E15-449
Cambridge, MA 02139 USA
cynthiab@media.mit.edu

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract
This study explores how people's social response toward a humanoid robot can change when we vary the number of the active degrees of freedom in the robot's head and face area. We investigate this problem by conducting two wizard-of-oz user studies that situate an elder person in a self-disclosure dialogue with a remotely operated robot. In our first study, we investigated the effect of expressive head gestures with a four-degree-of-freedom neck. In the second study we focused on the face where we investigated the effect of expressive eyebrow movement versus active gaze and eyelid movement. In the first study, we found that participants are willing to disclose more to the robot when the robot moved its neck in an expressive manner. In the second study, our data suggests a trend where gaze and expressive eyelid movement results in more disclosure over eyebrow movement

Keywords
Agents and Intelligent Systems, Elderly, User Studies, Robots.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Design, Experimentation, Human Factors

Introduction

In creating a sociable robot, the design of its head and face movement is important since these areas can evoke emotional interaction between robots and humans [1]. For instance, humanoid robots have been designed to display basic emotive expressions through moving the neck, eyelids, eyebrows, mouth, and nose [e.g., 5]. However, oftentimes in industry, designers want to minimize the number of degrees of freedom (DOF) used in the head and face to reduce manufacturing costs. In this paper, we investigate how the expressive degrees of freedom of a robot's head and face affect people's social responses toward it in a self-disclosure task. We chose to use a self-disclosure task because it is a well-known task in human communication studies, and the task has relevance to social robot applications in health and eldercare scenarios.

For instance, elders receive frequent medical care and communication between doctors, nurses, and patients is always important. However, patients sometimes fail to disclose important information to their doctors or nurses for concern of being held in less regard (e.g., too vulnerable, too negligent, too difficult, etc.). How might the expressive abilities of a robot impact people's willingness to disclose personal information to it? Might a patient be more likely to disclose personal information to a robot where "losing face" is perceived as being less at stake?

To date, a few human-robot interaction studies have examined the relationship between a robot's physical attributes and corresponding human responses and social judgments to the robot [e.g., 8]. In this study, we take a more systematic approach to investigate how specific expressive features of a robot contribute to social judgments – e.g., expressive head movements, active eye movements, or eyebrow movements. Our goal is to provide researchers with basic design guidelines to help inform the design tradeoff in the social judgment "bang" for the degree of freedom "buck."

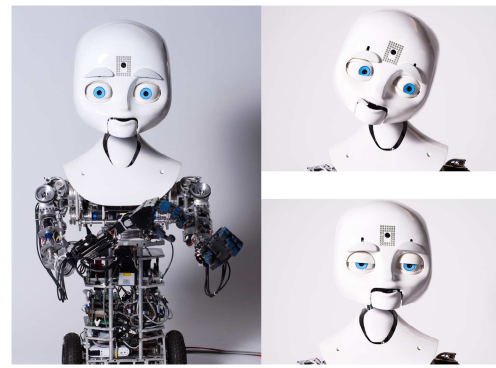

figure 1. The Mobile Dextrous and Social (MDS) robot (left). Two images on the right show examples of its facial expressions.

Mobile Social and Dexterous (MDS)

For the following experiments, we used the Mobile Dexterous and Social (MDS) robot in Figure 1. The robot combines mobility, dexterity, and social expressiveness with a highly articulate face for HRI research. The MDS is approximately 48 inches tall and has a 17 DOF face, a 4 DOF neck, two 4 DOF shoulders, two 4 DOF arm, and a 1 DOF hip rotate. The face joints include eyeballs (3 DOF for two eyes), eyelids (2 DOFs), eyebrows (2 DOFs), and a jaw (3 DOFs). The 4 DOF neck include head yaw, head roll, head pitch, and head forward and backward.

Self Disclosure

The self-disclosure task has been widely and successfully used in the field of communication, human computer interaction, human robot interaction, and behavioral economics to measure users or customers responses to different media [2,3,4,6]. In our self-disclosure task, the robot asks the human participant a number of questions of varying degrees of intimacy where some are more intrusive than others. Examples include questions about their family, romance, and most embarrassing or regretful moments. The robot is expressively animated during the interaction, and depending on the experimental condition, the robot can gesture with its arms, perform expressive head gestures, perform eyebrow movements, or move its eyeballs and eyelids.

Experiment Design

We conducted two different user studies, each following a between-subjects design. The first experiment explored the effect of adding expressive neck movements with a still face. The two conditions we compared are "arm-gestures-only" (gesture only), "arm-gestures-with-neck-movement" (gesture neck). Our hypothesis for this experiment is:

H1 Participants will be more likely to disclose more information when the neck DOFs are moving.

For the second experiment, three different face configurations were compared. In all cases, the robot performs expressive arm and head gestures. The three face conditions were: "eyebrows", "eyeballs", and "expressive eyeballs" (see Figure 2). The configurations were designed to test the effect of affect expressed through the eyebrows, versus attention cues expressed through gaze (a more cognitive/thinking cue), versus conveying both attention and affect in the eye region through active gaze and expressive eyelids. The more expressive eyes are considered to provide more human-like feelings for a robot than less-expressive eyes [2]. In the conditions where the robot's eyes or eyebrows did not move at all, we hid these features of the robot behind dark "sunglasses". We believed that showing "dead" but human-looking features would violate people's expectations toward how the robot ought to move these features. The hypothesis for the second study is:

H2 Participants will be more likely to reveal their information when the facial features are more expressive.

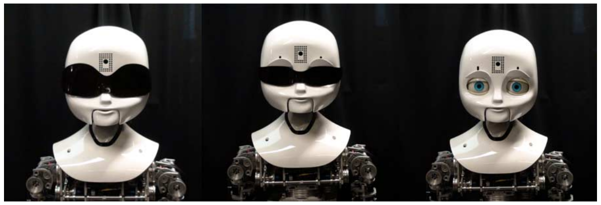

figure 2. Three face configurations. The first one was used for the "arm gestures only" and "arm gestures with neck" conditions. The second was used for the "eyebrows" condition and the last one was used for the two "eyeball" conditions.

Participants

We recruited 84 participants from senior centers around the Greater Boston Area. The participant's average age was 74. We put flyers on the walls inside the senior center to recruit subjects. As compensation for participating in the study, hey were given a choice of a mug a photo of the MDS robot or a $10 gift certificate from a local store.
In the first study there were n=15 participants in the "gestures only" condition (13 female and 2 male), and n=15 participatns in the "gesture neck" condition (11 female and 4 male).

In the second study, there were n=17 participants in the "eyebrows" condition (12 female and 5 male), n=13 in the "eyeballs" condition (10 female and 3 male), and n=17 in the "expressive eyeball" condition (11 female and 6 male).

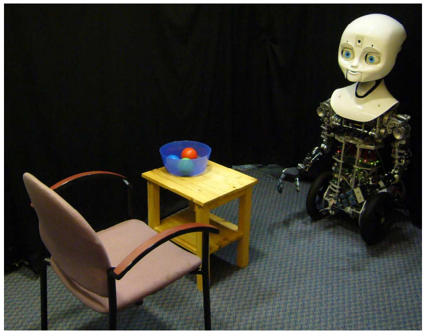

figure 3. The Study Setup. This setup was used in the elder study.

Procedure

Subjects were asked to participate in one-on-one interaction with the humanoid robot in a private room as shown in Figure 3. A person serving as a guide escorted the participant to this room. Once the guide left, the robot would begin the interaction by telling the participant a little about itself: its name, how it was been built, and a quick overview of its mobile, dexterous, and social abilities. The robot's name in the dialogue was Nexi and it had a young female's voice (pre-recorded by a female college student). As the robot explained about its "sensors", "arms", and "wheelbase", it would as the participant which part it should explain first to make the self-introduction a little interactive.

After its introduction, the robot asked the participant questions of varying degrees of intimacy. Before each question, the robot always provided its own short disclosure of personal information and commented about its own experience. The use of reciprocal comments are known to elicit more responses from participants [6].

The participant was made aware that he or she might stop the interaction at any time during the interaction. After they completed the given interaction, participants were asked to complete an exit survey.

Equipments

A private room was used for the study. Pipe-and-drape walls defined the study space (6 feet by 8 feet) inside the room as in Figure 3. The robot operator's desk was installed behind the curtain wall. Three desktop computers were used to run the software system and record video and audio of the interaction. The three FireFly MV cameras were installed around the walls to watch the robot, the participant, and the entire scene. Videos were streamed to the desktop computers to be viewed and recorded. The cameras were hidden behind the curtains so that only the lenses protruded. A boom microphone was attached the camera directed at the robot.

The robot operator sat behind the curtain wall and controlled the robot using the desktop computer. The operator could see all three video streams on the screen and hear the sounds being recorded using a headphone. The operator could control the robot using a Graphical User Interface (GUI). The operator clicked GUI buttons to invoke gestures and related vocal comments for each question. The operator waited for each participant to finish his/her response and initiated the next question. The operator could then choose comments following each response such as "I see", "I understand", "Yes", "Uh-huh", and "I am sorry to hear that" before starting the next question to smooth out the transitions between questions and to make the interaction seem more conversational.

Dependent Variables

There are a number of ways to measure self-disclosure behavior. For instance, measures can be divided into two categories: breadth and depth. The most widely used method is to count the number of words of the participant's utterance to understand the breadth of participants' responses [7]. The depth measure rates the intimacy of participants' responses [7]. The depth measures quantified the number of emotional adjectives (e.g., happy, sad) present in the participant's disclosures. This reflects how emotionally charged the participants were when answering each question.

Two human coders (both blind to the hypotheses) were used to score the data using these measures ($\alpha = 0.799$). In addition, we performed video analysis on the recorded videos to quantify the frequency of the participants smile. The coder was instructed to watch "Lip Corner Puller", "Cheek Raiser", "Lid Tightner" (AU12, AU6, AU7 in Ekman's Facial Action Coding System (FACS) [3]). The frequency of smile was calculated when AU12 was activated and when there was laughter present.

Once the participant left the private room, they were asked to fill out the questionnaire. The questionnaire was comprised of questions to ascertain the robot's perceived credibility, engagement, trust, and liking. In addition there were some questions that explored people's attitudes on the future use robots in society and general demographic information. Demographic information included gender, ethnicity, education, and technical knowledge.

Credibility was measured using D. K. Berlo's Source Credibility Scale ($\alpha = 0.731$) [see 4], and engagement was measured using Lombard and Ditton's scales ($\alpha = 0.887$) [see 6]. Trust was measured using 15 questions in a seven-point Likert scale ($\alpha = 0.834$) [see 9]. Liking was measured using 7 questions with a seven-point Likert scale asking how experienced, friendly, informed, intelligent, qualified, skilled, trained, understandable the robot was ($\alpha = 0.834$) [see 10].

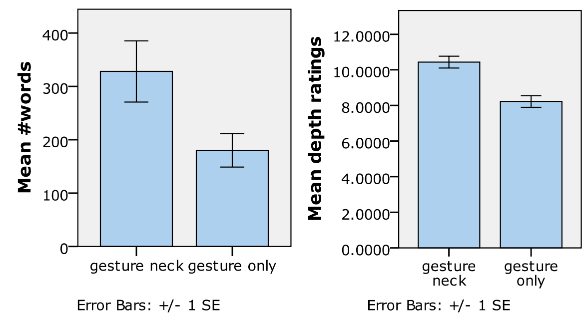

figure 4. The total number of words and the depth ratings in the first study (presented in means with standard errors).

Results and Discussions

For the first study, we tested the hypothesis (**H1**) using a 1-way ANOVA (Analysis of Variance). The total number of words, the coders' ratings on depth, the number of emotional words, the frequency of smiles were all significantly different in the two conditions. Our **H1** hypothesis was upheld: $F(1,28) = 5.114$, $p<.032$, $F(1,28) = 22.534$, $p<.01$, $F(1,28) = 5.586$, $p<.025$, and $F(1,28) = 3.936$, $p <.058$.

For the second study, we have preliminary results. To date, we have only examined the total number of words that participants spoke. Although we could see a trend of increase in the graph shown in Figure 5 (that hints at **H2**), we have not yet achieved statistical significance to show difference between conditions yet using 1-Way ANOVA: $F(2,45)= .947$, $p=.395$.

We have learned that it is significantly harder to manage quality in animations that are played back via the DOFs in the facial features. People are very sensitive to attentional cues, mutual gaze, and when eye contact is established and broken – all determine how a robot should move its eyes in a way that feels natural and does not violate people's expectations for how eyes ought to move (both in a biological sense and in a socially appropriate manner). We are fine-tuning the robot's gaze behavior and are planning to conduct an additional user study.

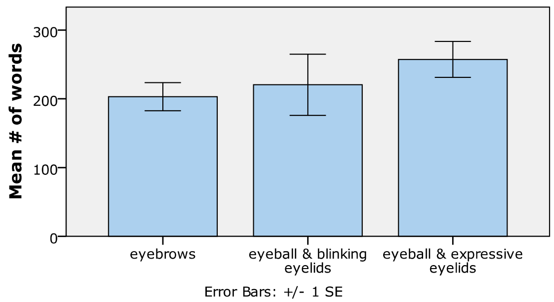

figure 5. The total number of words in the second study (presented in means with standard errors).

Conclusion

In the first study, our data supports the hypothesis that elderly people disclose with more willingness when the robot moves its head in an expressive manner. This is consistent with principles of classical animation, where the primary axis of motion conveys believability of character and contributes to liking and engagement. For the second study, we will continue to refine the robot's eye behavior and will conduct an additional study to investigate **H2**.

References

[1] Breazeal, C.L. *Designing Sociable Robots*. MIT Press, Cambridge, Mass, 2002.

[2] DiSalvo, C.F., Gemperle, F., Forlizzi, J., and Kiesler, S. All robots are not created equal: The design and perception of humanoid robot heads. *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, ACM (2002), 326.

[3] Ekman, P. and Friesen, W.V. *Facial action coding system: A technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto, CA, 1978.

[4] Fogg, B.J. and Tseng, H. The elements of computer credibility. *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, ACM New York, NY, USA (1999), 80-87.

[5] Fong, T., Nourbakhsh, I., and Dautenhahn, K.A survey of socially interactive robots. *Robotics and Autonomous Systems 42*, 3-4 (2003), 143-166.

[6] Lombard, M. and Ditton, T. At the heart of it all: The concept of presence. *Journal of computer-mediated communication 3*, 2 (1997), 20.

[7] Moon, Y. Intimate Exchanges: Using Computers to Elicit Self-Disclosure From Consumers. *Journal of Consumer Research 26*, 4 (2000), 323-339.

[8] Powers, A. and Kiesler, S. The advisor robot: tracing people's mental model from a robot's physical attributes. *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, ACM (2006), 218-225.

[9] Rubin, R.B., Palmgreen, P., Sypher, H.E., and Beatty, M. Communication research measures: A sourcebook. Lawrence Erlbaum, 2004.

[10] Takayama, L., Groom, V., and Nass, C. I'm sorry, Dave: I'm afraid I won't do that: social aspects of human-agent conflict. Proceedings of the 27th international conference on Human factors in computing systems, ACM New York, NY, USA (2009), 2099-2108.