Action Speaks Louder than Words:

Users’ Social Responses to Robots’ Movements and Voices

Paper Presented at the 18th Conference of the International Society for Presence Research

Kun Xu
Klein College of Media and Communication
Temple University
Email: Kun.xu@temple.edu
Abstract

This study focuses on two pairs of social cues of the social robot Alpha. It examines whether primary social cues are more likely than secondary social cues in evoking users’ attachment for the robot. The study suggests that a robot with gestural movements is more likeable than a robot with non-gestural movements. Users are more motivated to have future interaction with a robot with gestural movements than with non-gestural movements. Though there is no difference between a human voice and a synthetic voice in evoking users’ social responses, this study suggests that prior robot use experience moderates the relationship between the robot’s voices and social responses. For those who never interacted with robots before, a robot with a human voice would be more likely to lead users to experience medium-as-social-actor presence, perceive the robot to be attractive, and have future use intentions. For those who had experience using robots, a robot with synthetic voice would be more likely to lead to medium-as-social-actor presence, perceived attraction of the robot, and future use intentions.

Keywords: Medium-as-social-actor presence, the Computers are Social Actors paradigm, the Media are Social Actors paradigm, human-robot interaction, social responses
Action Speaks Louder than Words: Users’ Social Responses to Robots’ Movements and Voices

Introduction

In the past decade, communication scholars have witnessed a growing number of studies about social robots. A social robot is defined as “an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact” (Bartneck & Forlizzi, 2004, p. 592). Social robots have been applied in many areas including space travel, education, nursing, psychotherapy, and so on. For example, Bartneck, Reichenbach, and Carpenter (2008) designed social robots to take care of elderly people. The telepresence robot company Double (2017) designed its robots to connect homebound students to real classroom environments. Duquette, Michaud, and Mercier (2008) found that social robots can help autistic children focus their attention and express their emotions.

To improve the communication effectiveness between users and robots, one approach that designers can take is to increase the perceived attractiveness of the robot and increase users’ intention of future interaction with the robot. By prioritizing and optimizing some social cues that a robot can present to the users, designers can enhance the perceived socialness of the robot. Therefore, this preliminary study focuses on two social cues that a robot can manifest: Voice and Movement. This study seeks to understand whether a robot’s human voice would evoke stronger social responses than its synthetic voice and whether gestural movements would evoke stronger social responses than non-gestural movements. This study also examines the influence of users’ prior robot exposure on their social attitudes toward the robots.
Media are Social Actors

To study users’ social responses to robots, this study uses the Media are Social Actors (MASA) paradigm (Xu & Lombard, 2016). Based on the Computers are Social Actors (CASA) paradigm, the MASA paradigm suggests that users not only respond to computers as social actors, but also a wide range of media technologies including voice assistants, smartphones, televisions, and robots. In the MASA paradigm, Xu and Lombard (2016) argued that users may experience medium-as-social-actor presence when responding to the social cues of media technologies. Medium-as-social-actor presence refers to the idea that individuals perceive media technologies as social entities (ISPR, 2000, presence defined). The social cues of media technologies include voice, human face, language, interactivity, engagement, and so on (Fogg, 2002; Nass, 2004).

Drawing on the CASA paradigm, Xu and Lombard (2016) further distinguished primary social cues (e.g., human shape, human voice) and secondary social cues (e.g., movements, texts). Primary social cues are those that are “most salient and central to human’s perception of socialness” (p. 11), while secondary social cues are less salient and central and less likely to evoke users’ social responses to media technologies. Primary social cues should be more likely than secondary social cues in evoking users’ social responses. The MASA paradigm further indicates that it is not only the social cues, but also the combination of social cues, abstract human characteristics, individual factors, and contextual factors that lead technology users to perceive media as social actors. Therefore, this study focuses on the interaction between social cues and individual factors, especially how a robot’s social cues interact with users’ robot exposure in predicting medium-as-social-actor presence experience and other social reactions.
Specifically, two pairs of social cues are examined in the study. They are human voice versus synthetic voice and gestural movements versus non-gestural movements. Voice can be defined as “the sound produced by humans and other vertebrates using the lungs and the vocal folds in the larynx, or voice box” (NIH, 2017). Research has suggested that people are sensitive to human voices. For instance, Nass and colleagues (1997) used a male voice and a female voice to study users’ attribution of gender stereotypes to computers. Bracken and Lombard (2004) embedded a female voice to a computer in the educational setting and found that children demonstrated flattery effects when receiving the computer’s praise. In addition, Nass and Steuer (1993) suggested that human voice has special acoustic properties that can be easily recognized by humans.

Compared to human voice, synthetic voice can be understood as a computer-generated voice. It has been used in media technologies such as automatic telephone answer machines and subway station reporting systems. Unlike human voice, synthetic voice is perceived as unnatural and sometimes unpleasant (Gong & Lai, 2005). Though synthetic voice can be as effective as human voice in communicating the content of the messages, it reduces the effects of paralinguistic cues such as tones and accents (Nass & Lee, 2001).

Gestures and movements are kinetic cues that can potentially express social actors’ emotions, attitudes, and intentions. Compared to movements, gestures can be understood as human-like movements that deliver social meanings. For example, Bevan and Fraser (2015) found that those who shook hands with a social robot before negotiating with it were more likely to reach an agreement than those who did not shake hands with the social robot.

Humans also demonstrate social responses to the motion of objects even if these objects do not present human behavior patterns. Ju and Takayama (2009) found that humans can
perceive door movements to be social. The Heider and Simmel (1944) experiment showed that humans attribute social behavior to moving dots and interpret these dots as humans. These movements are considered as secondary social cues in the MASA paradigm. Compared to gestural movements, these non-gestural movements should be less likely to lead to medium-as-social-actor presence and users’ social reactions to media technologies.

In addition to the technology features, individuals may differ in their experiences of using robots. Prior research on the relationship between users’ computer use experience and their social attitudes toward computers has demonstrated inconsistent findings. For example, Nass and colleagues (1995) did not find a significant relationship between computer use experience and users’ attitudes toward computers. However, Johnson, Gardner, and Wiles (2004) found that only experienced computer users reported more trust in a computer when they were flattered by the computer. As it is unclear whether and how users’ technology use experience affects their social responses, this study investigates the influence of participants’ robot use experience on their social responses. Based on the review of literature, the following hypotheses and research questions are proposed.

**H1**: Compared to a synthetic voice, a social robot with a human voice will lead to greater levels of medium-as-social-actor presence, perceived attraction of the robot, and intention of future use.

**H2**: Compared to non-gestural movements, a social robot with gestural movements will lead to greater levels of medium-as-social-actor presence, perceived attraction of the robot, and intention of future use.
RQ1: How will a social robot’s kinetic cues interact with its vocal cues in predicting users’ medium-as-social-actor presence, perceived attraction of the robot, and intention of future use?

RQ2: How will robot use experience interact with the robot’s vocal cues and movement cues in predicting users’ medium-as-social-actor presence, perceived attraction of the robot, and intention of future use?

Method

Participants

A total of 110 participants were recruited from a public university in Northeast United States to voluntarily participate in an experiment. Participants received extra credits for their participation. They were also informed of the benefits and risks of the experiment. After data cleaning, 107 participants’ responses were included in the final analyses. Among them, 54 participants (50.5%) were males and 53 participants (49.5%) were females. Their age ranged from 18 to 34 years old.

Experiment Design

A social robot called “Alpha” designed by the UBTECH company was used as the experiment device. “Alpha” is a humanoid robot designed for social interaction. It is 15.67 inches tall, 8.19 inches wide, and 4.80 inches deep. There are 16 servomotor joints on its body and limbs, which allow Alpha to have 16 degrees of freedom. With these servomotor joints, the robot can be programmed to demonstrate different actions. The robot has a mono aural amplifier. It is installed with LED lights to indicate the status of the robot. The speaker is on the back of the robot.
The experiment used a 2 (human voice vs. synthetic voice) X 2 (gestures vs. non-gestural movements) between-subjects factorial design. Participants were randomly assigned to one of the four conditions: human voice with gestures, synthetic voice with gestures, human voice with non-gestural movements, and synthetic voice with non-gestural movements.

To create the human voice, a pre-recorded human-voiced message was installed into the robot through its programming software. A male voice rather than a female voice was used to match the appearance of the robot Alpha. To create the synthetic voice, a text-to-speech software was used to transform the message into a synthetic voice. The message in the synthetic voice conditions was the same as the message in the human voice conditions.

To create the gestures of the robot, two types of gestures from Krauss and colleagues’ (1996) typology of movements were used. They are symbolic gestures and conversational gestures. Symbolic gestures are the gestures that have specific cultural meanings without speech (e.g., thumbs-up, waving). Conversational gestures are the ones that occur at the same time with speech but do not necessarily have clear meanings without speech. Some conversational gestures (i.e., motor movements) are irrelevant to the linguistic content of the speech (Krauss et al., 1996). Examples of motor movements include spreading one’s hands or using hand movements in front of one’s chest. Other conversational gestures (i.e., lexical movements) are used to describe the linguistic content in the communicator’s speech (Krauss et al., 1996). For example, they can be used to point to a direction or depict the size of an object. A list of symbolic gestures and conversational gestures identified in prior research (Cabibihan et al., 2012; Hoffman & Ju, 2012; Salem et al., 2013) was designed into the gesture conditions. These gestures included hands on hips (i.e., to show aggression or confusion), hands on head (i.e., to show tiredness), opening arms (i.e., to show sincerity and openness), both hands behind head (i.e., to show
confidence or a relaxing state), waving hands (i.e., greetings), and bowing (i.e., to greet or express appreciation).

To create the non-gestural movements of the robot, a few random hand movements were used in the non-gestural movement conditions. These hand movements were designed not to have specific meanings. They did not include symbolic gestures or conversational gestures. The robot Alpha was programmed to hold hands flat, raise hands up, and stretch arms forward during the speech. The number of hand movements and the timing of making movements in the non-gestural movement conditions were designed to be same as those in the gesture conditions.

**Research Procedures**

When the participants entered the lab, they were first led to a laptop on the table. They were asked to read the consent form for the experiment and were informed of the benefits and risks of the experiment. After signing the consent form, they were randomly assigned to one of the four conditions of the experiment. To avoid the influence from other participants, only one person participated in the experiment at a time. The participants were first required to answer the questions about their demographic information (e.g., age, gender, etc.). Then they were led to another table where the robot Alpha was standing.

The robot Alpha was put about 25 inches away from the participant. The participants were told that the robot would give a brief self-introduction. They were not allowed to touch the robot. They were also advised not to move their seat position to control for the effects of distance between the participants and the robot.

Then the social robot Alpha began its two-minute self-introduction. The self-introduction includes four parts. In the first part, the robot introduced its name, where it was made, and its basic functions. In the second part, the robot introduced what it can do. In the third part, it briefly
talked about its communication experience with humans. In the last part, it talked about its potential applications. Throughout the whole process of self-introduction, speech and gestures were presented simultaneously. After observing Alpha’s self-introduction, the participants were asked to fill out more questionnaire items on the laptop.

**Measures**

The measure of medium-as-social-actor presence was adapted from the measures of social presence in the contexts of human-computer interaction (Lee & Nass, 2005) and human-robot interaction (Lee et al., 2005; Lee et al., 2006). The participants were asked to report their feelings on a Likert-type scale with six 10-point statements (1 = not at all, 10 = very much). The statements include “How much did you feel as if you were interacting with an intelligent being?” and “How much did you feel as if you and the robot Alpha were communicating with each other?”

The measure of perceived social attraction was adapted from the previous measures of social attraction (Jung & Lee, 2004; Lee et al., 2006; McCroskey & McCain, 1974). The participants were asked to report the extent of agreement on a 10-point Likert-type scale with four statements (1 = strongly disagree, 10 = strongly agree). The statements include “I could establish a personal relationship with the robot Alpha” and “I think I had a good time with the robot Alpha.”

The measure of intention of future use was adapted from the Shin and Choo (2011) measure of intention to use. The measure is a 10-point Likert-type scale with three items (1 = strongly disagree, 10 = strongly agree). Examples of the statements include “I would like to use a robot like Alpha again” and “I would be willing to talk to the robot Alpha in the future.”
Participants were also asked about demographic information including their age, gender, and so on. Robot exposure was measured by asking how many times participants had interacted with a humanoid robot in the past year. Based on the frequency of the responses, robot exposure was coded into a binary variable (0 = never, 1 = once or more than once).

Results

Three-way ANOVA was conducted to test the hypotheses and research questions. Results suggested that robot Alpha’s voice did not have main effects on users’ medium-as-social-actor presence, $F(1, 98) = 1.50, p > .05$. That is, the human voice ($M = 6.94, SD = 1.71$) did not evoke a greater level of medium-as-social-actor presence than the synthetic voice ($M = 6.71, SD = 1.87$). The robot’s movements did not have main effects on users’ medium-as-social-actor presence, $F(1, 98) = .06, p > .05$, which means that the robot’s gestures ($M = 6.88, SD = 1.96$) did not evoke a greater level of medium-as-social-actor presence than the robot’s non-gestural movements ($M = 6.77, SD = 1.60$). There was no interaction between the robot’s voice and movements in predicting users’ medium-as-social-actor presence, $F(1, 98) = .21, p > .05$. However, results further suggested that participants’ robot exposure interacted with the robot’s voice, $F(1, 98) = 8.11, p < .01$. Specifically, those who never interacted with a robot before perceived the robot with human voice as more socially present, while those who had prior experience in interacting with a robot perceived the robot Alpha with the human voice as less socially present (Figure 1).

Figure 1.
Results also suggested that robot Alpha’s voice did not have main effects on perceived attraction of the robot, $F(1, 98) = 1.91, p > .05$. That is, the human voice ($M = 5.87, SD = 2.44$) did not evoke a greater level of perceived attraction than the synthetic voice ($M = 5.43, SD = 2.58$). The robot’s movements had main effects on perceived attraction of the robot, $F(1, 98) = 7.83, p < .01$. The robot’s gestures ($M = 6.39, SD = 2.49$) evoked a greater level of perceived attraction than the robot’s non-gestural movements ($M = 4.88, SD = 2.31$). There was no interaction between the robot’s voice and movements in predicting perceived attraction of the robot, $F(1, 98) = .14, p > .05$. Results further suggested that participants’ robot exposure interacted with the robot’s voice, $F(1, 98) = 5.12, p < .05$. Those who never interacted with a robot before perceived the robot with human voice as more attractive, while those who had prior experience in interacting with a robot perceived the robot with the synthetic voice as more attractive (Figure 2).
Results suggested that robot Alpha’s voice did not have main effects on intention of future use, $F(1, 98) = 1.31, p > .05$. That is, the human voice ($M = 7.82, SD = 2.25$) did not evoke a greater level of future use intentions than the synthetic voice ($M = 7.64, SD = 2.32$). The robot’s movements had main effects on future use intentions, $F(1, 98) = 5.05, p < .05$. The robot’s gestures ($M = 8.23, SD = 2.24$) led to more intentions of future use than the robot’s non-gestural movements ($M = 7.21, SD = 2.23$). There was no interaction between the robot’s voice and movements in predicting intentions of future use, $F(1, 98) = .00, p > .05$. However, results further suggested that participants’ robot exposure interacted with the robot’s voice, $F(1, 98) = 6.02, p < .05$. Specifically, those who never interacted with a robot before had more intentions of using a robot with human voice, while those who had prior experience in interacting with a robot
has less intentions of future use of a robot with human voice (Figure 3). The main effects of the robot’s voices and movements have been showed in Table 1.

Figure 3.

*Interaction between robot use experience and robot voice on intention of future use*

![Interaction between robot use experience and robot voice on intention of future use](image)

Table 1

*The main effects of voices and movements*

<table>
<thead>
<tr>
<th></th>
<th>Human voice</th>
<th>Synthetic voice</th>
<th>Main effects</th>
<th>Gestural movements</th>
<th>Non-gestural movements</th>
<th>Main effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>F (df1, df2)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>F (df1, df2)</td>
</tr>
<tr>
<td>Medium-as-social-actor presence</td>
<td>6.94 (1.70)</td>
<td>6.71 (1.87)</td>
<td>1.50 (1, 98)</td>
<td>6.88 (1.96)</td>
<td>6.77 (1.60)</td>
<td>.06 (1, 98)</td>
</tr>
<tr>
<td>Perceived attraction</td>
<td>5.87 (2.44)</td>
<td>5.43 (2.58)</td>
<td>1.91 (1, 98)</td>
<td>6.40 (2.49)</td>
<td>4.88 (2.31)</td>
<td>7.83 (1, 98)**</td>
</tr>
<tr>
<td>Future use intention</td>
<td>7.82 (2.25)</td>
<td>7.64 (2.32)</td>
<td>1.31 (1, 98)</td>
<td>8.23 (2.34)</td>
<td>7.21 (2.23)</td>
<td>5.05 (1, 98)*</td>
</tr>
</tbody>
</table>

Note: *means p < .05, **means p < .01. M means the mean value. SD means standard deviation. Df means degrees of freedom.
Discussion and Conclusions

This study examines whether primary social cues are more likely than secondary social cues in evoking users’ attachment for a social robot. The study suggests that a robot with gestural movements is more likeable than a robot with non-gestural movements. Users are more motivated to have future interaction with a robot with gestural movements than with non-gestural movements. Though there is no difference between human voice and synthetic voice in evoking users’ social responses, this study suggests that prior robot use experience moderates the relationship between the robot’s voice and social responses. For those who never interacted with robots before, a robot with human voice would be more likely to lead users to experience medium-as-social-actor presence, perceive the robot to be attractive, and have future use intentions. For those who had experience using robots, a robot with synthetic voice would be more likely to lead to medium-as-social-actor presence, perceived attraction of the robot, and future use intentions.

The study has both theoretical and practical implications. First, it partially supports the propositions of the MASA paradigm where primary social cues are more likely to evoke social responses than secondary social cues. Compared to random movements, users are attracted to the robots that show human behavior and deliver social meanings. On the other side, the study shows that a human voice may not necessarily lead to greater levels of social responses than a synthetic voice. It may be because users would be more connected to the robots that show consistency in their appearances and faces. Prior study has corroborated the postulation. Gong and Nass (2007) found that when participants communicated with a social actor that had a human face and a synthetic voice or one that had a computer-generated human-like face and human voice, these participants were reluctant to disclose private information to the social actor. They perceived the
social actor as unconvincing. Only when the social actor on the computer screen had a human face and a human voice or when the social actor had a computer-generated face and a synthetic voice did participants report more trust in the social actor. Although the robot in the current study has a human shape, it still has many mechanical features. The consistency between these mechanical features and the synthetic voice may evoke users’ strong social responses.

The study implies that designers can take users’ prior robot use experience into consideration. The study suggests that those who had no prior robot use experience were sensitive to the human voice and would report strong medium-as-social-actor presence experience, while those who had experiences in using robots would report strong medium-as-social-actor presence when exposed to a machine voice. It may be because those who were experienced in interacting with robots would prefer consistency between a robot’s voice and its appearances. It may also imply that over time, users may want to avoid uncanny valley effects (Mori, MacDorman, & Kageki, 2012) and desire a robot that has fewer human features.

Overall, this study demonstrates a positive relationship between a robot’s gestural movements and users’ medium-as-social-actor presence of the social robot. Compared to the power of human gestures, the human voice did not show persuasive effects on users’ social responses to the robot. The effects of the human voice on medium-as-social-actor presence, perceived attractiveness of the robot, and intention of future use were contingent upon users’ robot exposure.
References


