Service Robot Feature and Interface Design Effects on User Emotional Responses

Tao Zhang, Biwen Zhu, Lashanda Hodge, Manida Swangnetr, Prithima Mosaly & David Kaber*

Edward P. Fitts Department of Industrial and Systems Engineering
North Carolina State University

Abstract

For service robots developed to assist nurses in routine patient services (e.g., medicine delivery), patient emotional experiences with the robot may be as important as robot task performance in terms of assessments of effectiveness and acceptance. The objective of this study was to understand the effect of different features of service robots on elderly people’s perceptions and emotional responses in a simulated medicine delivery task. Twenty four participants sat in a simulated patient room and a service robot entered and delivered a bag of “medicine” to them. Repeated trials were used to present variations on three robot features, including facial configuration, voice messaging and interactivity. Participant ratings of robot humanness (perceived anthropomorphism (PA)) were collected using a subjective rating scale along with emotional responses on arousal (excited – bored) and valence (happy – unhappy) using the self-assessment manikin (SAM) questionnaire after each trial. Results indicated the three types of robot features all promoted higher PA, arousal and valence, compared to a control condition (a robot configuration without any of the features). It was also found that the three types of features had different utility for stimulating participant emotional responses in terms of arousal and valence. In general, the results indicated that adding anthropomorphic features (face and voice) to service robots not only lead to higher perceptions of humanness, but also promote more positive emotional responses in elderly users. It is expected that results from this study could be used as a basis for developing affective robot interface design guidelines to promote user emotional experiences.

INTRODUCTION

Due to the current shortage of nurses for patient services (Picard & Klein, 2002), existing nursing staffs are overworked and there is a high potential for errors in task performance (Cleary et al., 2002; Goodin, 2003; Rogers et al., 2004). As the nursing shortage is expected to increase in the future, it is possible that service robots will need to take over some routine patient service tasks and communicate directly with patients. Currently, there are commercially available service robots (e.g., Aethon Tug) that can navigate autonomously in large-scale hospital workplaces and deliver medicines to nurse stations. However, these robots typically do not deliver medicines directly to patients and thus are not designed to support close interaction with patients.

If future service robots are to directly interact with patients like human nurses do, patient emotional reactions will be an important aspect of such interaction (Fong, Nourbakhsh, & Dautenhahn, 2003); however, few studies have examined emotional responses of users in interacting with robots (e.g., Kulic & Croft, 2006). In a healthcare environment, a patient’s emotional and social needs may be more important than traditional performance measures of human-robot interaction, such as efficiency and accuracy. Therefore, if knowledge of patient emotional responses to service robots is obtained, a design basis for future robots may be developed to facilitate positive patient experiences (e.g., in medicine delivery) and to increase the quality of healthcare.

In order to use emotional responses to assess patient-robot interaction (PRI), a method to identify specific emotion states is required. A psychological theory of emotion by Lang (1995) and Russell (1980) says that all emotional states can be located in a two-dimensional space, defined by arousal and valence. Russell (1980) described the typical course of an emotion as: (1) an event, or some combination of events (internal or external), producing a noticeable change in an individual’s state; (2) individual automatic perception of this change in their core affective state, in terms of arousal-sleep and pleasure-displeasure (also named as valence); and (3) individual labeling of the emotional state. Russell suggested that arousal and valence are the cognitive
dimensions of individual emotional states and are accessible to external measurement.

Related to user emotional experience with service robots, anthropomorphism is an important concept in robot interface design for motivating user interest and perceptions of robot capabilities. It is defined as “the act of attributing human-like qualities to non-human organisms or objects” (DiSalvo & Gemperle, 2003). Anthropomorphism in robot design may occur through physical design features or perceivable behaviors conveyed with physical movements or required user actions at an interface. Anthropomorphic features in robot design has been used to support social interaction and human emotional responses (Epley, Waytz, & Cacioppo, 2007; Fong et al., 2003). Epley et al. suggested that the perceived similarity of robots to humans not only influences the degree to which people anthropomorphize and interact with robots, but also increases a sense of social connectedness for users with robots. Based on Epley and colleagues theory, it can be expected that users would be more interested in interacting with a service robot and might have stronger social and emotional experiences with a robot with anthropomorphic features.

Based on review of existing robot technologies, there are three anthropomorphic features that appear to be important for direct interaction with patients. These include: presence or illusion of a human-like features – face/head (DiSalvo et al., 2002), voice capabilities (e.g., synthesized speech; Pollack et al., 2002) and intuitive methods for user interaction with robots (e.g., keypad or touch screen; Schaeffer & May, 1999). However, robot face features, voice and interactivity have not been tested to quantify their effects on patient perceptions and emotional responses. Related to this, there are also few emotion-oriented design guidelines for robot user interfaces.

In general, there is a need to empirically study how effective service robots can be in various interactive healthcare tasks. Zhang et al. (2008) recently defined a research framework for studying PRI (Figure 1). The framework was developed to motivate analyses of emotional responses as a measure of PRI, in general. On this basis, we sought to identify relationships between the different robot interfaces and patient emotional responses.

We examined the effect of robot facial configurations, voice messages and different levels of interactivity on user perceptions of arousal, valence and anthropomorphism in a typical healthcare task (medicine delivery). In order to accomplish this, a mobile robot platform was used (a PeopleBot). The platform allowed for flexibility, such that different interface configurations could be manipulated. The configurations were intended to represent different levels of anthropomorphism of the robot platform.

**METHOD**

**Participant**

Twenty-four participants (17 females and 7 males) residing at local senior centers were recruited for this study. They ranged in age from 63 to 91 years with a mean of 80.5 years and a standard deviation of 8.8 years. Nine participants said they knew about robots through books, movies or television. Two participants said they were familiar with the mechanics of robots, but none of the 24 participants had any direct experience in interacting with robots.

**Apparatus**

The PeopleBot platform was used for this experiment (see Figure 2). The platform has autonomous navigation capability, including map-based route programming and obstacle avoidance. The robot was integrated with additional equipment to introduce the anthropomorphic features (i.e., faces, voice messaging, and direct user interactivity). A set of mini cameras and a face mask were used to present different facial configurations. A tablet PC (HP tx2000)
was mounted on top of the PeopleBot and used to store and playback voice messages (WAV files) to users. The tablet was also used to present a user control interface in particular experiment trials.

**Figure 2. PeopleBot.**

**Scenario and Independent Variable Settings**

The experiment trials involved the robot performing a simulated medicine delivery task for participants. The robot entered a simulated patient room from a predefined starting position, with a container of medicine (empty) in its gripper. Subsequently, the robot traveled approximately 15ft at a speed of 1ft/s towards the participant seated on a couch. The robot stopped directly in front of the participant.

In each test trial the PeopleBot platform was augmented with one of the anthropomorphic features. Prior research has demonstrated that the selected features significantly affect people’s expectations of service robot capability (Fong et al., 2003). The face was either an abstract or human-like face. Figures 3a and 3b present these two configurations. Voice messaging refers to the capability of the robot to present speech during the delivery task. Either synthesized or digitized (human) voice messages were used. Interactivity refers to whether the user is required to visually or physically attend to the robot during the interaction. Depending on the experiment condition, participants either simply read a visual message on the tablet PC (Figure 4) or they pressed a button on the touch screen to confirm the delivery before the medicine was released.

Under all other conditions, participants simply waited for the robot to open the gripper so they could take the medicine. After a participant retrieved the medicine, the robot waited for a short period of time, and then turned around and moved out of the room. The average length of a trial was approximately 3 minutes. Each participant completed a total of 14 trials (= (1 control condition + 6 test conditions) × 2 replications).

**Figure 3a. Abstract face configuration.**

**Figure 3b. Human-like face configuration.**

**Figure 4. Robot visual message on touch screen.**

**Experiment Design and Dependent Variables**

A single factor, randomized complete block design was used in this study. As mentioned above, each participant was exposed to the same robot condition twice. The experiment design did not include combinations of the various robot features due to limitations of the PeopleBot platform (e.g., the face and tablet PC could not be mounted on the robot at the
same time).

As dependent variables, two subjective measures were collected in this study, including: (1) perceived anthropomorphism (PA); and (2) ratings of emotional state, in terms of arousal and valence. Perceived anthropomorphism was defined as the perception of humanness based on both the robot physical shape and interaction experience. We developed a PA questionnaire, based on Catrambone, Stasko and Xiao’s (2002) research. The questionnaire addressed four dimensions of anthropomorphism in service robot design, including physical appearance, expressiveness, task handling, and user subjective experience. Participants initially made pair-wise comparisons of the dimensions in terms of importance for characterizing the degree of humanness of a robot. Weights were determined for each factor based on the pair-wise comparisons. Participants then rated the various robot configurations along the dimensions after each test trial. A total PA score was calculated as a weighted sum of participant ratings of each factor. With respect to the ratings of arousal and valence, the self-assessment manikin (SAM) questionnaire (Bradley & Lang, 1994) was used. It consisted of two rows of cartoon characters, representing bored to excited and happy to unhappy expressions (corresponding to arousal and valence). The participants were to circle those characters that best represented their emotional experience. The range of PA score was between 0 and 100 and the SAM ratings were between 1 and 9.

Hypotheses

We assessed the effect of the different anthropomorphic features and the feature settings on PA and the subjective ratings of emotion, including arousal and valence. Relative to this assessment, we formulated two hypotheses:

Hypothesis 1: It was expected that the different types of robot features would lead to different user perceptions of robot humanness measured by the PA questionnaire. For each robot feature, as the level progressed from the control condition (no feature) to abstract human representation (face mask with cameras; digitized speech for voice messaging; or touch screen confirmation of the medicine delivery), user responses to perceived anthropomorphism were expected to increase.

Hypothesis 2: Related the Hypothesis 1, it was expected that different robot interface features would lead to different emotional responses, captured in terms of valence and arousal ratings with the SAM questionnaire (Bradley & Lang, 1994). Similar to Hypothesis 1, as each interface feature progressed from “none” to “abstract human” to “human-like”, user valence and arousal ratings were expected to shift towards happiness and excited, accordingly.

Procedure

After participants completed and signed an informed consent form, a brief introduction on the concept of a service robot and applications in healthcare tasks were presented. A background information survey was conducted that included questions regarding the participant’s age, gender, occupation (if any), level of familiarity with robots, and prior experience with robots (at work, as toys, in movies/books, in TV shows, in museums or in schools). In addition, the level of technical knowledge of robots was also collected (Dautenhahn et al., 2005). A training session was provided in which participants were presented with video clips and pictures of several existing service robots and they were introduced to the PA and SAM questionnaires.

Prior to formal testing, participants were told that a robot would deliver a prescribed medicine to them and they needed to accept the simulated medicine container (empty) from the robot. At the end of each trial, participants were asked to complete the SAM questionnaire and give ratings regarding the four factors of the PA measure. After participants completed the 14 test trials, a final interview was conducted in which they were asked to provide their impressions on the robot configurations, in general, and the types of features they thought were important for considering a robot to be human-like.
Data Analysis

The subjective rating data for PA, arousal and valence was transformed to z-scores for each participant in advance of statistical analysis. This was done to account for individual differences in internal scaling of emotions and PA.

Between-feature analysis: A series of one-way (feature type) ANOVAs were conducted on the subjective measures of arousal, valence and PA. The analyses were used to identify any significant features in perceptions of humanness of the robot configurations or emotional states.

Within-feature analysis: In addition to assessing the effects of the types of features, we also conducted ANOVAs to assess the effect of various levels of each feature on PA, arousal and valence ratings. Post-hoc analyses were conducted on all significant features using Duncan’s Multiple Range test. An alpha value of 0.05 was used to determine statistical significance across analysis methods.

RESULT

Between-feature Analysis

ANOVA results revealed significant effects of feature type on PA, arousal and valence ratings (Table 1). Post hoc analysis using Duncan’s test indicated that all feature types (face, voice message and interactivity) significantly differed from the control condition in terms of PA, arousal and valence (Figure 5). The three types of robot features all promoted higher PA, arousal and valence, compared to the control condition. However, there were no significant differences among the three features in terms of influence on the subjective responses.

<table>
<thead>
<tr>
<th>Factor</th>
<th>PA</th>
<th>Arousal</th>
<th>Valence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F(3, 309) = 9.46, p &lt; 0.0001</td>
<td>F(3, 309) = 4.85, p &lt; 0.0001</td>
<td>F(3, 309) = 14.61, p &lt; 0.0001</td>
</tr>
</tbody>
</table>

Table 1. ANOVA results on PA, arousal and valence for feature type.

Within-feature Analysis

ANOVA results revealed that all levels of face, voice and interactivity significantly differed in effect on PA, arousal and valence (see Table 2). Furthermore, post hoc analysis using Duncan’s test indicated that certain feature levels were significantly different from others. Arousal ratings were higher for face and voice features, as compared to the control condition. Arousal ratings for interactivity increased from the control (none) to visual message and user confirmation. For valence and PA, ratings were significantly higher for the voice and interactivity features, as compared to the control condition. PA ratings for face also increased from the control (none) to abstract and human-like face. On a feature basis, the human-like face yielded higher scores for PA and valence than abstract face, but no significant difference was seen for arousal (Figure 6). There was no significant difference in PA, arousal and valence between synthesized and digitized (human) voice (Figure 7). For interactivity, the confirmation condition resulted in a higher score on arousal than visual message only condition, but not on PA and valence (Figure 8).

<table>
<thead>
<tr>
<th>Effect</th>
<th>PA</th>
<th>Arousal</th>
<th>Valence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F(2,118) = 9.46, p = 0.0002</td>
<td>F(2,118) = 4.85, p = 0.0095</td>
<td>F(2,118) = 14.61, p &lt; 0.0001</td>
</tr>
<tr>
<td>Face</td>
<td>9.46</td>
<td>4.85</td>
<td>14.61</td>
</tr>
<tr>
<td>Voice</td>
<td>18.04</td>
<td>10.33</td>
<td>13.52</td>
</tr>
<tr>
<td>Interactivity</td>
<td>14.83</td>
<td>13.65</td>
<td>13.77</td>
</tr>
</tbody>
</table>

Table 2. ANOVA results on PA, arousal and valence within feature types.
Validation of Subjective Measures

Regarding the PA questionnaire, for the ratings on the four dimensions of perceived anthropomorphism (physical appearance, expressiveness, task handling, and user subjective experience), the overall standardized Cronbach's coefficient alpha was 0.723 for the first replication and 0.777 for the second replication. These two numbers are greater than the suggested value of 0.70 given by Nunnally and Bernstein (1994). This suggests that the PA questionnaire used in this study is a reliable measure of participant perception of robot humanness.

In order to determine whether the observed differences in PA and emotion ratings among the various feature settings were significant in the presence of individual differences, the mean square (MS) for the participant term, included in the ANOVA models, was used in additional F-tests on the specific feature types. The test results are summarized in Table 4 and suggest that the robot configuration manipulations were significant across participants. The results also indicate that the subjective measures were sensitive to differences in participant perceptions across experiment conditions.

<table>
<thead>
<tr>
<th>Subjective Measure</th>
<th>Effect</th>
<th>F(2, 23)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA</td>
<td>Face</td>
<td>16.93</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Interactivity</td>
<td>21.82</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Voice</td>
<td>23.09</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Arousal</td>
<td>Face</td>
<td>5.36</td>
<td>0.0123</td>
</tr>
<tr>
<td></td>
<td>Interactivity</td>
<td>17.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Voice</td>
<td>10.98</td>
<td>0.0004</td>
</tr>
<tr>
<td>Valence</td>
<td>Face</td>
<td>13.46</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Interactivity</td>
<td>31.60</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Voice</td>
<td>16.71</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Interview Results

In the final interview, we asked participants for their general impressions of the use of robot technology in health care environments. Fifteen (62.5%) of the participants said they would like to have delivery robots working in hospitals. However, they were concerned about robot accuracy in delivering the correct medicine to patients and in verifying that a patient actually took the medicine. The remaining participants either didn’t like the robot configurations or thought they were not ready for hospital use.

When we asked for comments on the specific robot features, 15 (62.5%) of the participants liked the voice interactivity, particularly the human voice. They also pointed out that the screen or display interactivity
was clearer in terms of delivering the message to the user and would be very useful for people with hearing problems. Five (20.8%) of the participants either didn’t like the face configurations or thought a face didn’t make a substantial difference in PA; however, they did prefer the human-like face to the abstract face.

**DISCUSSION**

In line with our expectation, the results of the between-feature comparison indicated that adding human-like features (face, voice and interactivity) to the robot interface promoted positive emotional responses. This is also in agreement with predictions from Epley et al. theory that users will have stronger emotional experiences with a robot with anthropomorphic features. However, it was not clear which type of feature would be most effective for conveying anthropomorphism and facilitating emotionally-rich interaction for patients. A possible reason for this is that users might consider all three types of features to be important to representing humanness (Disalvo et al., 2002) and thus their reactions to robot configurations with any type of human-like features were relatively consistent.

From the within-feature comparisons, the effect of the various levels of robot features on PA was consistent with our Hypothesis 1. When features became more human-like, participant ratings of PA increased. Disalvo et al. (2002) and Nass et al. (2001) made similar findings for face and voice features, respectively.

More importantly, it was found that the pattern of participant emotional response to the three types of features was different. This may be largely due to the resulting differences in user interaction supported by specific features. For example, abstract and human-like faces differed in valence ratings but not arousal. This is probably due to the fact that the face feature of the robot was static (no change or movement during the delivery). Therefore, participants merely observed the face when the robot approached them and this was the basis for the experience of valence (i.e., whether the robot face was pleasurable). On the contrary, the levels of interactivity (visual message and user confirmation) caused different arousal responses. Participants either read the information displayed on the tablet PC screen or used a stylus to click on the touch screen when required. Watching the robot respond to commands was more exciting for participants than passively receiving the medicine without knowing what the robot would do next. Therefore, interaction with the robot was more effective in promoting participant arousal. The lack of significant differences among the synthesized and digitized speech conditions may have been due to condition similarities. The length of the WAV files for both conditions was relatively short (7-8s) and the synthesized speech was generated by a state-of-art text-to-speech software applications. Several participants commented in the final interview that they could not differentiate the two voices (synthesized and recorded human).

**CONCLUSION**

Considering increases in healthcare service demands and the current shortage of nurses, service robots represent a technological solution for addressing routine patient service tasks. In such applications, there may be direct interaction of patients with robots and human emotions may play a critical role in the effectiveness of the interaction. Following this, we studied the effect of different robot interface features on human perceptions of robot humanness as well as emotional responses in a simulated healthcare task (medicine delivery). Our findings indicate that for elderly users, adding anthropomorphic features (face, speech and interactivity) to service robots leads to perceptions of humanness and positive emotional experiences. Within each feature type, perception of humanness and positive emotional responses increase with increasing feature complexity or realism (compared to the human). Furthermore, humanoid robot features differ in terms of their economy for driving participant emotional responses. These results provide a basis for determining appropriate implementations of robot features, including face, voice and interactivity, to promote accurate user perceptions of robot interactive capability and positive user emotion experiences. Additional studies are needed to generate a collection of affective robot interface design guidelines.
ACKNOWLEDGMENTS

This research was supported by the Department of Industrial & Systems Engineering at North Carolina State University and conducted in the Cognitive Ergonomics Laboratory.

REFERENCES


