Interlibrary Loans and Journal Article Requests

Notice Warning Concerning Copyright Restrictions:

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted materials.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One specified condition is that the photocopy or reproduction is not to be *"used for any purpose other than private study, scholarship, or research."* If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

Upon receipt of this reproduction of the publication you have requested, you understand that the publication may be protected by copyright law. You also understand that you are expected to comply with copyright law and to limit your use to one for private study, scholarship, or research and not to systematically reproduce or in any way make available multiple copies of the publication.

The Stephen B. Thacker CDC Library reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Terms and Conditions for items sent by e-mail:

The contents of the attached document may be protected by copyright law. The CDC copyright policy outlines the responsibilities and guidance related to the reproduction of copyrighted materials at CDC. If the document is protected by copyright law, the following restrictions apply:

- You may print only one paper copy, from which you may not make further copies, except as maybe allowed by law.
- You may not make further electronic copies or convert the file into any other format.
- You may not cut and paste or otherwise alter the text.

Intel Serv Robotics (2010) 3:73–88
DOI 10.1007/s11370-010-0060-9

ORIGINAL RESEARCH PAPER

Service robot feature design effects on user perceptions and emotional responses

Tao Zhang · David B. Kaber · Biwen Zhu · Manida Swangnetr · Prithima Mosaly · Lashanda Hodge

Received: 12 July 2009 / Accepted: 8 February 2010 / Published online: 27 February 2010

Abstract Service robots have been developed to assist nurses in routine patient services. Prior research has recognized that patient emotional experiences with robots may be as important as robot task performance in terms of user acceptance and assessments of effectiveness. The objective of this study was to understand the effect of different service robot interface features on elderly perceptions and emotional responses in a simulated medicine delivery task. Twenty-four participants sat in a simulated patient room and a service robot 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 heart rate (HR) and galvanic skin response (GSR) were collected. Participant ratings of robot humanness [perceived anthropomorphism (PA)] were collected post-trial along with subjective ratings of arousal (bored–excited) and valence (unhappy–happy) using the self-assessment manikin (SAM) questionnaire. Results indicated the presence of all three types of robot features promoted higher PA, arousal and valence, compared to a control condition (a robot without any of the features). Participant physiological responses varied with events in their interaction with the robot. The three types of features also had different utility for stimulating participant arousal and valence, as well as physiological responses. In general, results indicated that adding anthropomorphic and interactive features to service robots promoted positive emotional responses [increased excitement (GSR) and happiness (HR)] 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.

Keywords Service robots · Anthropomorphism · Emotion · Physiological variables · Fuzzy modeling

T. Zhang
Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, USA

D. B. Kaber (✉) · B. Zhu · M. Swangnetr · L. Hodge
Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, USA
e-mail: dbkaber@ncsu.edu

P. Mosaly
Industrial Extension Services, North Carolina State University, Raleigh, USA

1 Introduction

With an aging population and ever increasing number of patients in hospitals, there is a high demand for patient services. Unfortunately, there has been a nursing staff shortage among healthcare providers in the US since 1998 [1]. Existing nursing staffs are overworked and there is a high potential for errors in task performance [2–4]. As the nursing shortage is expected to increase in the future, it is possible that service robots will be used to perform some routine patient service tasks and communicate directly with patients. Currently, there are commercially available robots (e.g., Aethon Tug) that can navigate autonomously in large-scale hospital workplaces and deliver medicines from a pharmacy 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.

Future service robots may be expected to interact with patients (especially elderly with healthcare needs) and patient emotional responses will be an important aspect of such interaction. Based on previous research on believable agents, Goetz et al. [5] suggested that in order for robotic assistants to be effective, they should exhibit naturalistic behaviours and appropriate emotions, and should require little or no learning or effort on the part of the user. People have the natural tendency to treat computers, communication technologies and interactive systems with emotional ties as in human-human social interaction [6]. This type of emotional tie is important for effective human-robot interaction, particularly in service contexts. Some previous research has focused on developing robots capable of affective expressions (e.g., [7]) or object delivery mechanisms between users and robots through manipulators [8–10]; however, few empirical studies have been conducted to assess emotional responses of users in interacting with robots (e.g., [11]). In a healthcare environment, measures of patients' emotional and social needs may complement traditional performance measures of human-robot interaction (HRI), such as efficiency and accuracy [12], in modelling and designing successful patient–robot interaction scenarios. Robots causing positive emotional experiences for users are expected to support more pleasurable and productive collaborations with humans. Therefore, there is a need to understand potential patient emotional responses to service robots and to provide a design basis for future robots to facilitate positive emotions and effective patient–robot interaction.

1.1 Measures of emotional responses

Using emotional responses to evaluate patient–robot interaction requires an effective technique of identifying specific emotion states within an emotional space. Three types of emotion models exist in the psychology literature, including: basic emotion theory [13,14], dimensional emotion theory [15,16], and models from appraisal theory (e.g., [17]). Basic emotion theory contends that there is a concise set of primary emotions (anger, disgust, fear, happiness, sadness, and surprise; see [18]) and that they are distinguishable from each other and other affective phenomena [19]. Dimensional emotion theory argues that all emotional states can be located in a two-dimensional space, defined by *arousal* and *valence*. Russell [16] described the typical course of an emotion as: (1) an event (internal or external) producing a change in an individual's state; (2) individual perception of the change in their core affective state, in terms of *arousal-sleep* and *pleasure-displeasure* (also referred to as *valence*); and (3) individual labelling of the emotion state. Russell suggested that arousal and valence are the cognitive dimensions of individual emotion states and are accessible to external measurement. Appraisal theory emphasizes that emotions are elicited by evaluations (*appraisals*) of events and situations, and differentiated by a set of appraisal criteria [20,21], including: novelty and expectancy, relevance and goal conduciveness, agency or responsibility, and perceived control, power or coping potential [22].

The above emotion models lead to different approaches of observing and assessing emotions in practice, based on either subjective or objective measures. Some practical issues in emotion measurement include the fact that mixed emotions are more likely to be evoked in many human–human interactions (and HRI) rather than a single, pure emotion [18]. Furthermore, in a HRI context, the stimuli for potential emotions (e.g., participant verbal expressions and body language) may vary less than in human–human interaction. Consequently, humans in interacting with robots may not experience the same range of distinguishable emotions as they experience in their daily lives in interacting with other humans. The continuous nature of emotional experiences in this applied setting may also not be captured by discrete measures of emotion state. Subjective measures based on dimensional emotion theory, such as the Affect Grid [23] and the Self-Assessment Manikin [24], allow for quick assessments of user emotional experiences but they may aggregate responses over the course of many events.

In addition to subjective measures, there is evidence suggesting that human physiology relates to emotional states or events [25,26]. Physiological variables found to be correlated with various emotions include: heart rate (HR), blood pressure, respiration, electrodermal activity (EDA) and galvanic skin response (GSR), as well as facial EMG (Electromyography). These measures can complement subjective measures by identifying emotional experiences for specific events and that subjects fail to reflect in subjective responses, either because they could not recall or they considered the feature to be unimportant [27]. Although physiological measures can be noisy and difficult to interpret, leading to lower specificity in emotion state identification, there is on-going research in this area. Recently, Mandryk and Atkins [28] presented a method for continuously classifying emotional states of computer/video game users based on a set of physiological variables and the use of fuzzy logic as well as the dimensional emotion model. They developed a fuzzy logic model to transform HR, GSR and facial EMG (smiling and frowning activity) into arousal and valence states. A second fuzzy logic model was then used to transform arousal and valence classifications into five lower-level emotion states related to the gaming situation, including: boredom, challenge, excitement, frustration and fun. Results from the fuzzy logic models revealed the same trends as self-reported emotions for fun, boredom and excitement. The advantage of the fuzzy logic approach is that the emotions of a user can be quantitatively assessed from physiological data during an entire test trial; thus, revealing variance within a complete emotional experience.

1.2 Anthropomorphism in service robot interface design

Related to emotional experience with service robots, anthropomorphism is an important concept in robot interface design for motivating user interest and perceptions of robot capabilities. By definition, anthropomorphism is user attribution of human-like qualities to non-human organisms or objects, particularly computers and robots [29–32]. It is also a design strategy to implement human-like features on a non-human object (e.g., the Kismet robot designed by Breazeal [33]). Anthropomorphism in robot design may occur through physical design features or required user actions at an interface. That is, some anthropomorphic features may be implicit and users must interact with an interface to fully understand these features. For example, a user might talk with a robot using a natural language dialogue interface in order to learn that the robot has emotions, personality, or other human-like characteristics. Although there is an on-going discussion over whether or not to use anthropomorphic robot designs, in general, it has been suggested that the use of human-like characteristics in robot design may support the sense of social interaction with robots [34,35]. That is, people might interact with robots in a manner similar to how they interact with other humans, as a result of perceptions of anthropomorphic cues [36]. For example, Bruce, Nourbakhsh and Simmons [37] found that having an expressive face and indicating robot attention to the user through movement made a robot more compelling and inviting to interact with. Further, anthropomorphic characteristics may contribute to the elicitation of robot acceptance and human compliance with robot behaviours (e.g., [5,8]), which is particularly important in patient service tasks like medicine delivery. In summary, it can be expected that users would be more interested in interacting with a service robot with appropriate designed anthropomorphic features, and this interaction, by its social nature, could not only promote positive emotional experiences for users with robots, but also lead to increased ease of use, user compliance and thus effectiveness of the human–robot interaction.

1.3 Motivation and objective

In general, there is a need to empirically study how effective service robots can be in interactive healthcare tasks. Based on review of existing robot technologies [38], there are three anthropomorphic features that appear to be important for direct interaction with patients, including (1) presence or illusion of a human-like features such as face or head [39], (2) voice capabilities (e.g., synthesized speech [40]), and (3) intuitive methods for user interaction with a robot (e.g., keypad or touch screen [41]). Although there have been studies of objective user responses (e.g., reaction time) to different robot feature manipulations [42–45], these features have not been tested for effects on user emotions. Related to this, there are few emotion-oriented design guidelines for robot interfaces. The objective of this study was to examine the effect of service robot anthropomorphic features (face configuration, voice messaging, and user interactivity) on user perceptions and emotional experiences based on first impressions in a typical patient service task (medicine delivery).

2 Methodology

2.1 Participants

Twenty-four participants (17 females and 7 males) at two local senior centers were recruited for the experiment. They ranged in age from 64 to 91 years ($M = 80.5$, $SD = 8.8$). Nine participants said they knew about robots through books, movies or television. Two participants said they had some knowledge about the mechanics of robots, but none of the participants had any direct experience in interacting with robots.

2.2 Apparatus

A PeopleBot platform (Fig. 1) was used for this experiment. The robot has autonomous navigation capability, including

Fig. 1 PeopleBot platform (also the control condition)

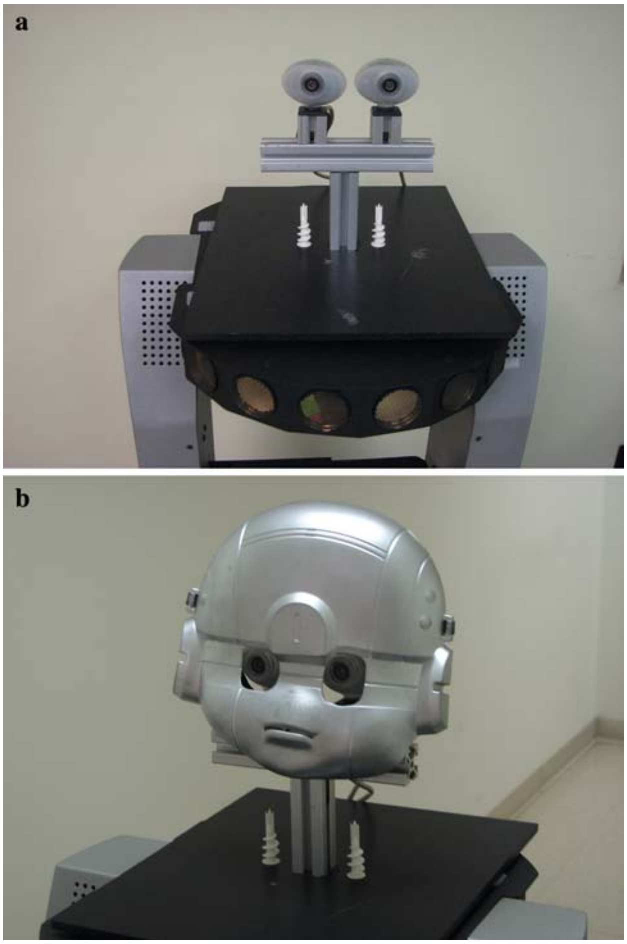


Fig. 2 **a** Abstract face configuration. **b** Human-like face configuration

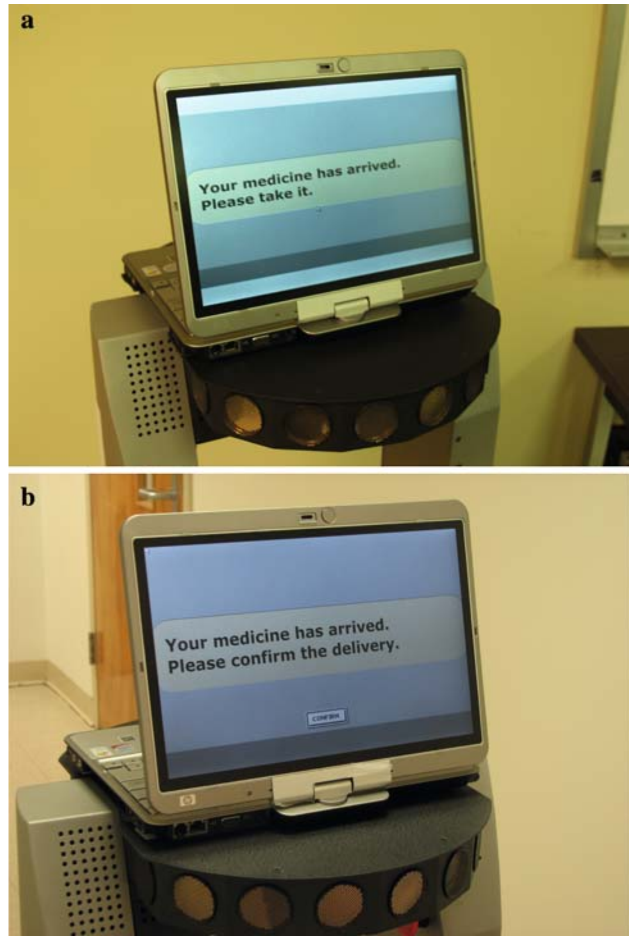


Fig. 3 **a** Visual message from robot on touch screen. **b** Visual message and confirmation from robot on touch screen

map-based route programming and obstacle avoidance. It was integrated with additional equipment, including a set of mini cameras and a face mask 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 participants. The tablet PC was also used to present a user control interface (for interactivity) in particular experiment trials.

2.3 Independent variables

The experiment trials involved the robot performing a simulated medicine delivery task for participants. In each trial, the PeopleBot was augmented with one of the anthropomorphic features (face, voice messaging, or user interactivity). There were two settings for each feature. The face was either an abstract or human-like face (see Fig. 2a, b). Either synthesized or digitized (human) voice messages with the same content were used during the delivery task. The physical appearance of the robot under both voice messaging conditions was the same as the control condition (Fig. 1). Interactivity refers to whether the participant was required to visually or physically attend to the robot during the interaction. In test trials, participants either simply read a visual message on the tablet PC or they pressed a button on the touch screen to confirm medicine delivery (see Fig. 3a, b). As mentioned, there was also a control condition in which the PeopleBot did not present any of the above features. Each participant completed a total of 14 trials [=(1 control condition + 6 test conditions) ×2 replications].

2.4 Response measures and types of analyses

Participant perceptions of the robot features were measured using a perceived anthropomorphism (PA) questionnaire. Perceived anthropomorphism was defined as the perception of humanness based on both the physical appearance of the robot and user interaction experience. We developed a questionnaire, based on Catrambone, Stasko and Xiao's [46] research, including four dimensions of anthropomorphism in service robot design: physical appearance, expressiveness,

Table 1 The dimensions of perceived anthropomorphism questionnaire

Dimension	Explanation
Physical appearance	The degree to which the robot physically looks like a human. Determinants include: roundness of shape, a "big" body, an erect posture, etc.
Expressiveness	The degree to which the robot could clearly present its state. Facial expressions and speech quality are two determinants.
Task handling	The degree to which the robot could perform intelligent assistance and take the initiative to start interaction with a user.
User subjective experience	The degree to which a user enjoys interacting with the robot and thinks that the robot is smart.

Table 2 Rules for fuzzy inference model

1.	If (GSR is high) then (Arousal is high)
2.	If (GSR is medium) then (Arousal is medium)
3.	If (GSR is low) then (Arousal is low)
4.	If (HR is low) then (Arousal is low)
5.	If (HR is high) then (Arousal is high)
6.	If (HR is high) and (GSR is low) then (Arousal is medium)
7.	If (HR is low) and (GSR is high) then (Arousal is medium)
8.	If (HR is medium) and (GSR is high) then (Arousal is high)
9.	If (HR is medium) and (GSR is medium) then (Arousal is medium)
10.	If (HR is low) then (Valence is low)
11.	If (HR is high) then (Valence is high)
12.	If (HR is medium) then (Valence is medium)

task handling, and user subjective experience (see Table 1 for definitions). 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 these comparisons. Participants then rated the various robot configurations along the dimensions after each trial. A total PA score was calculated as a weighted sum of participant ratings across factors. The range of PA scores was between 0 and 100. The relationship between different levels of anthropomorphic features and PA was later investigated by power function fitting.

Participant emotional responses to the interaction with the service robot were assessed with both subjective and physiological measures. Ratings of arousal and valence were made with the self-assessment manikin (SAM) questionnaire [24]. It consisted of two rows of cartoon characters, representing excited to bored and happy to unhappy expressions (corresponding to arousal and valence). The participants were to circle those characters that best represented their emotional experience. The SAM ratings were between 1 (bored and unhappy) and 9 (excited and happy).

Participant HR was measured using a Polar S810i heart rate monitor. Galvanic skin response (GSR) was collected through sensors attached to the palmar surface of the distal segment on participants' index and ring fingers of the non-dominant hand. To reduce signal noise, participants were asked not to move the hand with GSR sensors during test trials. The sampling frequency was set at 30 Hz according to suggestions by Dawson et al. [47].

Correlations among the subjective measures of emotion and physiological measures were examined. A fuzzy inference model adopted from Mandryk and Atkins's [28] study was also applied to infer participant emotional states in terms of arousal and valence from the physiological measures. The fuzzy inference rules are listed in Table 2. Results from the fuzzy inferences were further validated by ratings of arousal and valence from the SAM questionnaire.

2.5 Hypotheses

It was expected that the different types of robot features would lead to different user perceptions of robot humanness, emotional responses in terms of SAM ratings and physiological measures [Hypothesis (H)1]. For each robot feature, as the level progressed from the control condition (no feature) to abstract human representation (cameras for eyes; synthesized speech for voice messaging; or a visual delivery confirmation display) and then to a human-like representation (face mask with cameras; digitized speech for voice messaging; or touch screen confirmation of the medicine delivery), user responses to PA and SAM questionnaires, and HR and GSR measures were all expected to increase. Related to this, it was also expected that user responses to the different types of robot features would vary due to the characteristics of each feature type.

As perceived anthropomorphism is a measure of perceived salience (intensity and meaningfulness) of human-like features in the robot design, we also hypothesized (H2) that PA would operate as a power function of variations in the settings of robot features. This hypothesis was essentially an extension of Stevens' law, which states that the magnitude of sensation is simply the intensity of the physical stimulus raised to some power [48]. Specifically, the hypothesized relationship can be represented as,

$$\psi(I) = I^a.$$

Here, $\psi(I)$ is the PA score; I is the quantitative characteristic of robot features; and a is the exponent (power) that varies for different features. In this function, the perceived humanness elicited by a particular robot feature depends on the value of the exponent.

It was also hypothesized (H3) that there would be correlations between subjective measures of arousal and valence using the SAM and the physiological measures (HR and GSR). Previously, Bradley et al. [49] reported correlations between HR and valence with r-values ranging from 0.5 to 0.76 and a correlation between GSR and arousal with $r = 0.81$. Detenber et al. [50] also found a positive correlation between arousal and skin conductance level, but could not exactly replicate Bradley's findings on correlations between HR and valence. They did find a decrease in HR for negative valence stimuli, but the same occurred for positive stimuli with less strength. It appears that the relationships between GSR and arousal, and HR and valence are not definite. There is a need to validate these correlations in a HRI context.

Finally, with the possibility of correlations between SAM ratings and physiological measures, the fuzzy inference model integrating the physiological variable inputs was expected to yield outputs correlated with participant ratings of arousal and valence. It was expected that the fuzzy model could account for complex relations between the physiological measures and subjective perceptions, which might be ignored by simple linear correlation analysis.

2.6 Experiment design

A single factor, randomized complete block design was used in this study. The design did not include combinations of the various robot features due to limitations of the PeopleBot platform to support additional equipment (e.g., the face and the tablet PC could not be mounted on the robot at the same time). The sequence of robot conditions for each participant was randomized.

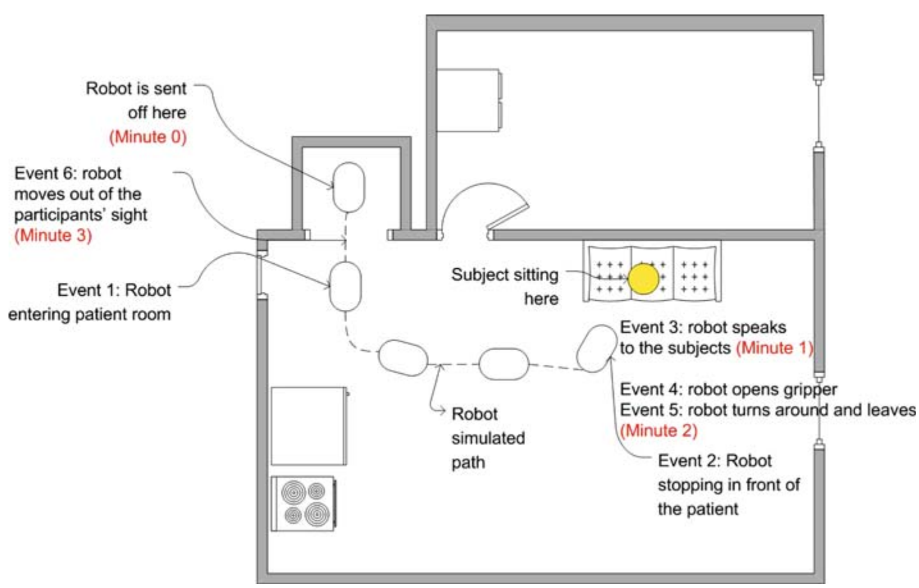
2.7 Procedure

Participants were initially briefed on the objectives of the study. After they read and signed an informed consent form, a background information survey was conducted that included questions on age, gender, occupation (if any), level of familiarity with robots, prior experience with robots (at work, as toys, in movie/books, in TV shows, in museums or in schools), and the level of technical knowledge of robots. Following the survey, a training session was provided in which participants were presented with video clips and photos of several existing service robots (e.g., Aethon Tug) and their applications in healthcare tasks. Participants were then asked to make pair-wise comparisons of the four dimensions of the PA questionnaire based on their perceptions of the service robots, in general. They were also introduced to the SAM questionnaire.

After the training session, the experimenter helped participants put on the chest strap of the heart rate monitor and GSR sensors. Participants were asked to relax and not to pay attention to these devices while their normal (baseline) HR and GSR responses were collected for 1 min. Participants continued wearing the HR chest strap and GSR sensors throughout the test trials.

Figure 4 presents the layout of the simulated patient room with five defined events along the robot movement path. At the beginning of each test trial, the robot entered the simulated patient room from a predefined position, holding in

Fig. 4 Layout of the simulated patient room and events along the robot path



its gripper an ordinary medicine bag with an empty container inside. The robot travelled approximately 15 ft across the room at a speed of 1 ft/s towards the participant seated on a couch. Previous studies on user preferences during a mobile robot delivery task [51,52] suggested that users tended to prefer a robot to approach them from either their left or right side, rather than straight at them. Therefore, in the test trials, the robot approached participants from their right side, turned and then stopped in front of them. The stopping point of the robot varied little relative to the participant position from trial-to-trial. Participants were told that a robot would deliver a prescribed medicine to them and that they needed to accept the medicine bag from the robot. A Velcro system, instead of a complex manipulator (e.g., [9]), was used to ensure the medicine bag did not fall to the floor when the robot opened its gripper. The robot was programmed to open the gripper when it reached the destination point. Participants did not need to stand up to grab the medicine bag, because the height of the robot gripper was well within their reach while sitting. Related to this, after participants sat down on the couch, the GSR sensors were attached to them so their movement was limited due to the length of the sensor connection. In the actual experiment, we did not observe large ranges of movement of participants in response to the location of the robot. The robot waited for the participant to take the bag for 30 seconds and then left the room. In actual experiment trials, participants took no more than 10 seconds to take the medicine bag. At the end of each trial, participants were asked to complete the SAM questionnaire and provide ratings on the four dimensions of the PA measure. The average length of a test trial was about 3 min. After participants completed the 14 test trials, a final interview was conducted in which they were asked to provide general impressions on the robot configurations and the types of features they thought were important for considering a service robot to be human-like.

3 Results

3.1 Descriptive statistics

The subjective ratings of PA, arousal, and valence were transformed to z-scores for each participant in advance of statistical analysis. This was done to account for individual differences in internal scaling of PA and emotions. Participant HR and GSR data were first normalized for each participant by calculating the proportional change relative to the baseline or at rest condition. As physiological measures (HR and GSR) are generally analyzed on an event-basis, several important events were identified in trials (see also Fig. 4), including: (1) the robot entering the simulated patient room; (2) the robot stopping in front of the participant; (3) the robot presenting voice messages to the participant (in trials with voice message conditions); (4) the robot opening its gripper and the participant taking the medicine container; (5) the robot turning and leaving the participant; and (6) the robot moving out of the participant's sight. Regarding the event time window for the GSR and HR responses, Lang et al. [53] used the largest value of skin conductance response (SCR) from 0.9 and 4.0 s and the fastest HR within the first 3 s after event onset. Levenson [54], citing Ekman [55], said that emotion typically lasts between 0.5 and 4 s. Dawson et al. [47] also said that any SCR beginning between 1 and 4 s, following a stimulus onset is considered to be elicited by that stimulus. Therefore, a 4-s time window after every event was used in the present study for the physiological data analysis. Within each time window, the maximum GSR value was identified and used as an indicator of phasic changes due to the event stimulus, and the average of HR was also calculated. The descriptive statistics for all response variables (in transformed units) across participants are presented in Table 3.

Table 3 Descriptive statistics of response variables (numbers in parentheses are standard deviations)

Feature type	Level	PA (z-score)	Arousal (z-score)	Valence (z-score)	HR (event average)	GSR (event maximum)
Face	Abstract	−0.2860 (0.989)	−0.125 (0.702)	−0.3 (0.839)	0.045 (0.058)	0.104 (0.138)
	Human-like	0.194 (0.95)	−0.017 (0.957)	0.298 (0.94)	0.044 (0.053)	0.085 (0.137)
	Overall	0.1179 (0.909)	0.092 (0.859)	0.105 (0.901)	0.044 (0.056)	0.095 (0.138)
Voice	Synthesized	0.178 (0.816)	0.036 (0.906)	0.167 (0.812)	0.046 (0.057)	0.101 (0.121)
	Digitized	0.363 (0.932)	0.28 (0.894)	0.197 (0.996)	0.047 (0.063)	0.094 (0.172)
	Overall	0.117 (0.909)	0.092 (0.86)	0.106 (0.901)	0.047 (0.060)	0.098 (0.148)
Interactivity	Visual message	0.184 (0.881)	−0.003 (0.837)	−0.011 (0.822)	0.047 (0.057)	0.088 (0.092)
	Confirmation	0.073 (0.782)	0.377 (0.764)	0.285 (0.879)	0.063 (0.061)	0.093 (0.209)
	Overall	0.118 (0.910)	0.092 (0.860)	0.106 (0.901)	0.055 (0.059)	0.090 (0.162)
Control		−0.708 (0.994)	−0.553 (0.804)	−0.635 (0.966)	0.042 (0.048)	0.104 (0.107)

Table 4 ANOVA results on PA, arousal and valence for feature type

Response variable	PA	Arousal	Valence
(Effect = Feature type)	$F(3, 309) = 12.5$, $p < 0.0001$, $\omega^2 = 0.4709$.	$F(3, 309) = 8.58$, $p < 0.0001$, $\omega^2 = 0.3697$.	$F(3, 309) = 9.59$, $p < 0.0001$, $\omega^2 = 0.3993$.

ω^2 (omega squared) is the effect size of the independent variable

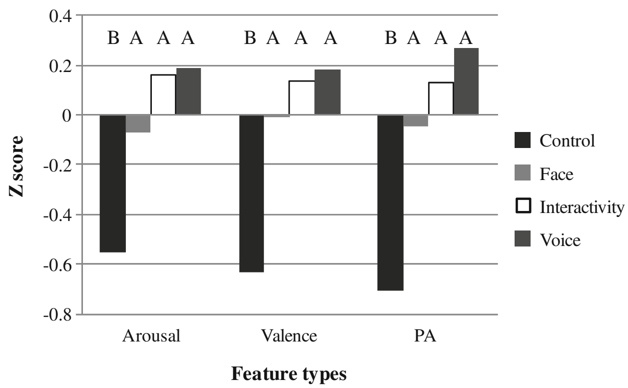


Fig. 5 Post hoc results on PA, arousal and valence for all feature types. (*Letters* indicate Duncan's means breakout results. Means with the same letter are not significantly different)

3.2 ANOVA modeling

3.2.1 Between-feature analyses

A series of one-way (feature type) ANOVAs were conducted on the subjective measures of arousal, valence and PA. The analyses were conducted to identify significant differences among features in terms of perceptions of humanness of the robot configurations or emotion states. Results revealed significant effects of feature type on all subjective measures ($F(3, 309) = 12.5, p < 0.0001$ for PA; $F(3, 309) = 8.58, p < 0.0001$ for arousal; $F(3, 309) = 9.59, p < 0.0001$ for valence; see also Table 4). 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 (Fig. 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 strength of influence on the subjective responses.

With respect to physiological measures, the HR data for two participants revealed variances substantially different from all other participants, possibly due to sensor errors. The data for these two participants was removed from the set and the HR responses for the rest of the sample exhibited normality and homogeneity of variance; that is, the degree of variability of the response measure appeared constant across the robot conditions. Consequently, an ANOVA was performed to test the significance of robot features and the levels of each feature. For the GSR response, because of ANOVA assumption violations, a nonparametric analysis (Friedman's Test) was conducted on all data.

A series of one-way ANOVAs were conducted to test the effect of feature type on HR during the events identified earlier. For Events 2, 4 and 5, the effect of feature type was significant (Table 5). Post hoc analysis using Duncan's test revealed that for Events 2 (the robot stopped in front of the participant) and 4 (the robot opened its gripper), HR was higher for the interactivity feature than voice, face and the control condition; HR for voice was also higher than face and the control condition; and there was no significant difference between voice and face, or face and the control condition (Fig. 6a, b). During Event 5 (the robot left the participant), interactivity and voice features produced higher HR than the face and control conditions (Fig. 7).

Friedman's test results indicated that participant GSR during Event 2 was significantly different across feature types (Table 6). Multiple comparisons showed that the interactivity condition produced significantly higher GSR than the face and voice conditions, and GSR under the control condition was actually higher than for the face condition (Table 7).

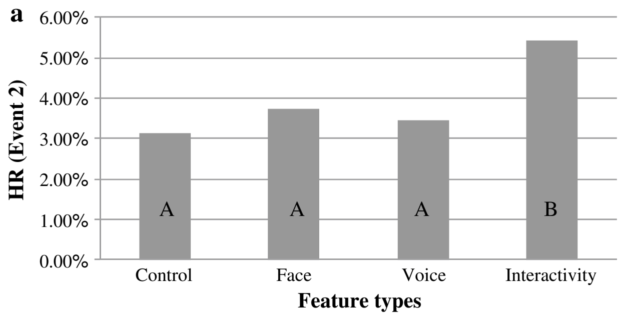
3.2.2 Within-feature analyses

In addition to assessing the effect of the type of robot feature, we also conducted ANOVAs to assess the effect of the

Table 5 ANOVA results on HR for feature type

HR	Event 1	Event 2	Event 4	Event 5	Event 6
(Effect = Feature type)	$F(3, 1367) = 1.84$, $p = 0.138$, $\omega^2 = 0.0143$	$F(3, 1326) = 11.84$, $p < 0.0001$, $\omega^2 = 0.1615$	$F(3, 1399) = 7.18$, $p < 0.0001$, $\omega^2 = 0.0943$	$F(3, 1374) = 9.95$, $p < 0.0001$, $\omega^2 = 0.1331$	$F(3, 1153) = 2.00$, $p = 0.112$, $\omega^2 = 0.0200$

ω^2 (omega squared) is the effect size of the independent variable



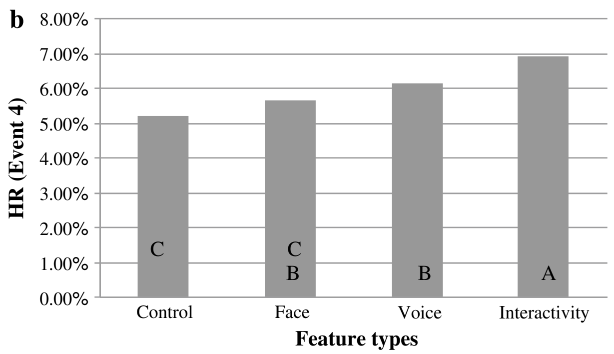


Fig. 6 **a** Post hoc result on Event 2 HR for feature types (*Letters* indicate Duncan's means breakout results. Means with the same letter are not significantly different.) **b** Post hoc result on Event 4 HR for feature types (*Letters* indicate Duncan's means breakout results. Means with the same letter are not significantly different)

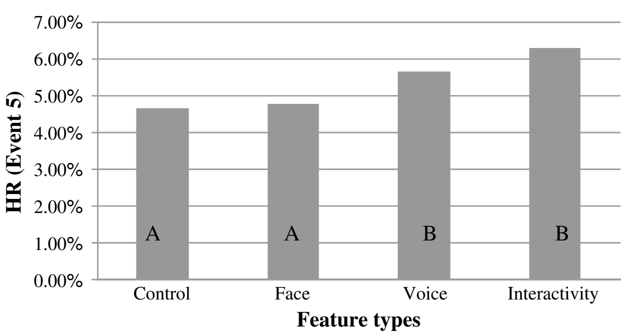


Fig. 7 Post hoc result on Event 5 HR for feature types (*Letters* indicate Duncan's means breakout results. Means with the same letter are not significantly different)

specific settings of each feature on the subjective response measures. Results revealed that all levels of face, voice and interactivity differed in effect on PA, arousal and valence (Table 8). Post hoc analysis using Duncan's test indicated that

Table 6 Friedman tests result on event GSR for feature type

GSR	Event 1	Event 2	Event 4	Event 5	Event 6
Effect = Feature type					
$\chi^2(3)$	2.65	10.85	5.75	1.25	5.4
p	0.449	0.012	0.124	0.741	0.144

Table 7 Multiple comparison results of Event 2 GSR for feature type

Comparison	$\chi^2(2)$	p	Conclusion
Control versus Face	4.167	0.041	Control > Face
Face versus Interactivity	8.166	0.004	Face < Interactivity
Voice versus Interactivity	4.167	0.04	Voice < Interactivity

certain feature levels were significantly different from others. Arousal ratings were higher for face and voice features, as compared to the control condition. They also increased from no interactivity (the control) to visual message confirmation and touch-screen 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 (no face) to abstract and human-like face.

On a feature basis, the human-like face yielded higher scores for PA and valence than the abstract face, but no significant differences were observed for arousal (Fig. 8). There was no significant difference in PA, arousal and valence between the synthesized and digitized (human) voices (Fig. 9). For interactivity, the touch-screen confirmation resulted in a higher score for arousal than visual messaging, but not on PA and valence (Fig. 10).

The results of ANOVAs on HR for each event are listed in Table 9. Post hoc analysis using Duncan's test revealed that the abstract face led to significantly higher HR compared to the human face for Event 6 (the robot moving out of participant sight). For the voice feature, the control condition lead to higher HR than synthesized voice for Event 1, and digitized voice increased HR compared to the control during Event 4. For the interactivity feature, touch-screen confirmation led to a significantly higher HR than the visual message and control conditions during Events 2, 4 and 6. In Event 5, HR for the touch-screen confirmation and visual message conditions was higher than the control condition.

Friedman's test for the effect of feature settings on GSR indicated that participant responses during Event 6 were significantly different among the levels of interactivity ($\chi^2(2) = 9.08$, $p = 0.011$). Multiple comparisons revealed higher GSR for visual messaging and touch-screen confirmation than the control condition. However, there was no significant difference in GSR between visual messaging and touch-screen trials (Table 10).

3.2.3 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

Table 8 ANOVA results on PA, arousal and valence within feature types

Effect	PA			Arousal			Valence		
	$F(2,118)$	p	ω^2	$F(2,118)$	p	ω^2	$F(2,118)$	p	ω^2
Face level	9.46	0.0002	0.5949	4.85	0.0095	0.4006	14.61	<0.0001	0.7026
Voice level	18.04	<0.0001	0.7474	10.33	<0.0001	0.6183	13.52	<0.0001	0.6849
Interactivity level	14.83	< 0.0001	0.7060	13.65	<0.0001	0.6871	13.77	<0.0001	0.6892

ω^2 (omega squared) is the effect size of the independent variable

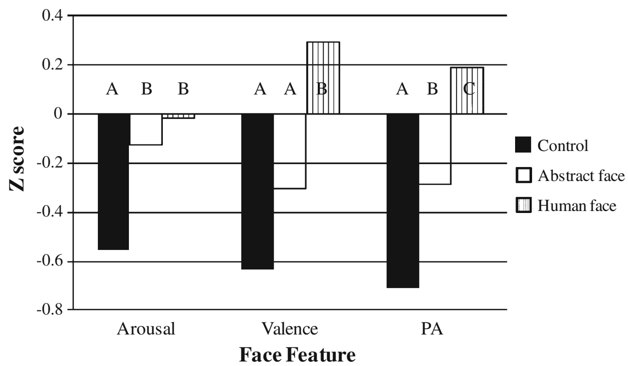


Fig. 8 Post hoc results for PA, arousal and valence with different face types (*Letters* indicate Duncan's means breakout results. Means with the same letter are not significantly different)

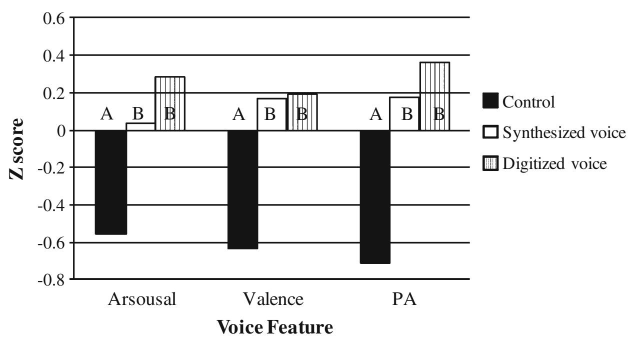


Fig. 9 Post hoc results for PA, arousal and valence with different voice types (*Letters* indicate Duncan's means breakout results. Means with the same letter are not significantly different)

coefficient (a measure of the reliability of ratings across subjects and conditions; α) 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 [56]. 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

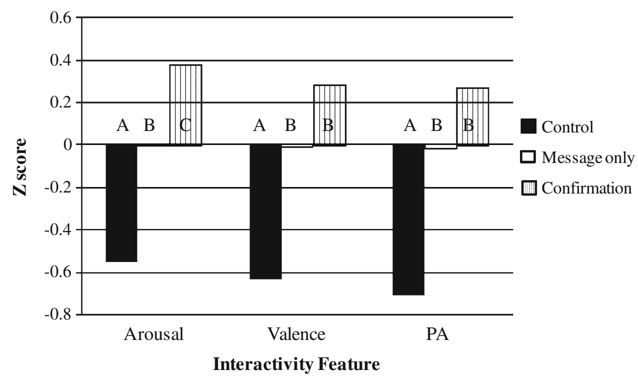


Fig. 10 Post hoc results for PA, arousal and valence with different interactivity types (*Letters* indicate Duncan's means breakout results. Means with the same letter are not significantly different)

Table 11 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.

3.3 Power function fitting

To quantify the various settings of robot features for fitting the PA ratings with a power function, we analyzed the physical characteristics of each feature. The number of perceivable facial characteristics (e.g., eyes, mouth, ears), the variance of the frequency spectrum of the voice messages, and the number of actions required of the user and robot during the medicine delivery were used to quantitatively represent the salience of the face, voice message and interactivity features, accordingly.

The power function fitting results are shown in the Table 12. The exponent values indicate the strength of influence of each human-like feature on PA. A Kruskal–Wallis test on the absolute values of the coefficients revealed a significant effect of feature type ($\chi^2 = 6.5340$, $p = 0.0381$). Multiple comparisons between feature types revealed that the coefficients for voice (across subjects) were significantly smaller than for face (Table 13).

Table 9 ANOVA results on HR within each feature type

Event 1	Event 2	Event 3	Event 4	Event 5	Event 6
Face					
$F(2, 584) = 0.48$, $p = 0.622$	$F(2, 559) = 0.59$, $p = 0.554$		$F(2592) = 0.61$, p = 0.546	$F(2, 583) = 0.23$, p = 0.794	$F(2, 490) = 3.04$, p = 0.049
Interaction					
$F(2, 578) = 0.59$, $p = 0.55$	$F(2, 578) = 21.10$, $p < 0.0001$, $\omega^2 = 0.4344$		$F(2, 599) = 23.64$, $p < 0.0001$, $\omega^2 = 0.4553$	$F(2, 580) = 8.84$, $p = 0.0002$, $\omega^2 = 0.2299$	$F(2, 477) = 3.14$, $p = 0.034$, $\omega^2 = 0.0895$
Voice					
$F(2, 583) = 3.43$, $p = 0.033$, $\omega^2 = 0.0843$	$F(2, 551) = 0.53$, $p = 0.59$	$F(1, 320) = 1.54$, $p = 0.216$	$F(2, 596) = 4.04$, $p = 0.018$, $\omega^2 = 0.1013$	$F(2, 591) = 2.27$, $p = 0.104$	$F(2, 504) = 2.39$, $p = 0.093$

ω^2 (omega squared) is the effect size of the independent variable

Table 10 Multiple comparison results of Event 6 GSR for interactivity level

Comparison	$\chi^2(2)$	p	Conclusion
Visual message versus Control	6	0.0143	Visual message > Control
Visual message versus Confirmation	0.16	0.683	Visual message = Confirmation
Confirmation versus Control	6	0.0143	Confirmation > Control

Table 11 F-tests using the MS for participant as an error term

Subjective measure	Effect	$F(2, 23)$	p
PA	Face	16.93	<0.0001
	Voice	23.09	<0.0001
	Interactivity	21.82	<0.0001
Arousal	Face	5.36	0.0123
	Voice	10.98	0.0004
	Interactivity	17.71	<0.0001
Valence	Face	13.46	0.0001
	Voice	16.71	<0.0001
	Interactivity	31.60	<0.0001

Table 12 Coefficients of power function fitting for each feature type

	Face	Voice	Interactivity
Exponent (Mean and SD for all 24 participants)	6.575 (4.988)	2.361 (2.358)	5.8669 (4.783)

Table 13 Results of multiple comparisons on coefficients for feature type

Coefficient comparison	Statistic	Results
Face versus voice	$W = -3.4325$	a(Face) > a(Voice)
Face versus interactivity	$W = -0.38$	a(Face) = a(Interactivity)
Voice versus interactivity	$W = 3.012$	a(Voice) = a(Interactivity)

The criterion at $\alpha = 0.05$ level is $W^* = 3.314$; any $|W| > W^*$ is significant

3.4 Fuzzy modeling

A diagrammatic representation of the fuzzy model used in this study is presented in Fig. 11. The model transformed the physiological response data (HR and GSR) to arousal and valence states based on defined fuzzy inference rules. The average HR and maximum GSR for each trial were inputs to the model and the outputs were fuzzy-inferred arousal and valence values for each trial.

The histograms for the normalized HR and GSR responses (across participants) are shown in Fig. 12a, b below. (Data on the two participants with large variances in HR were excluded.) The fuzzy membership functions for HR and GSR were determined based on the general shapes of the response histograms. Similar to Mandryk and Atkins [28], the mean response ± one standard deviation and the mean ± two standard deviations were used as the dividing points for the "low", "medium" and "high" ranges of HR and GSR data. The levels of the membership functions for arousal and valence outputs ("low", "medium" and "high") were defined to account for uniform proportions of the responses in the range from 0 to 100. Based on the inference rules in Table 2, the response surfaces for the fuzzy models on arousal and valence are shown in Fig. 13a, b.

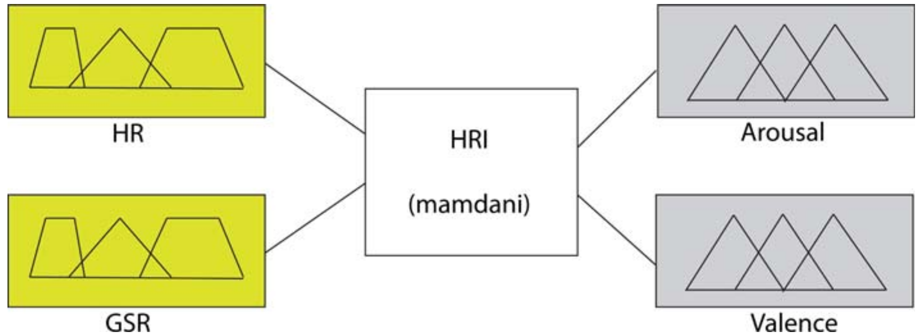


Fig. 11 Diagram of fuzzy logic model

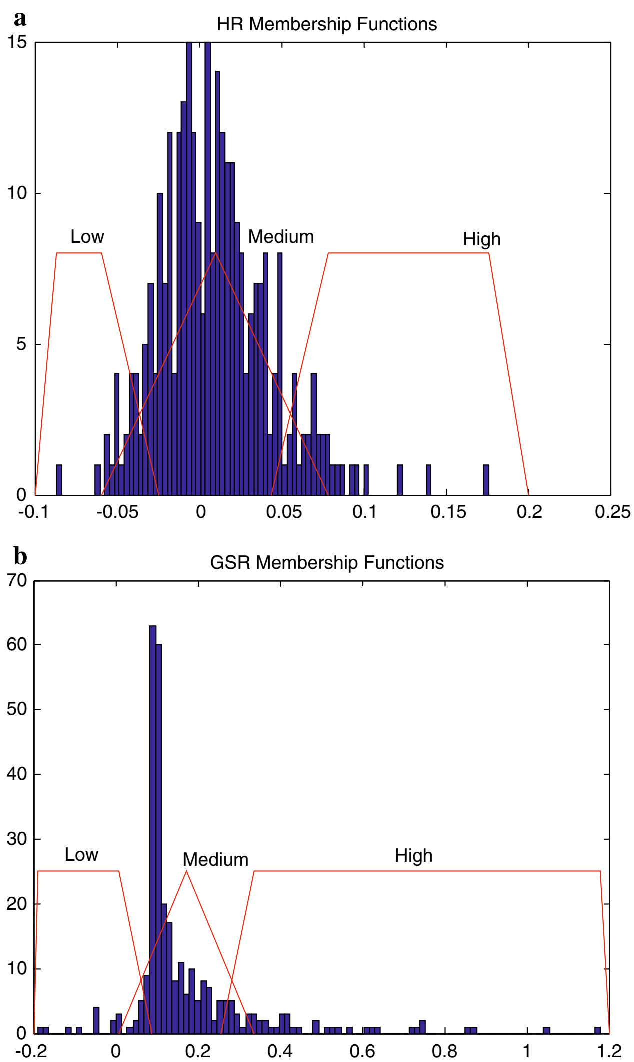


Fig. 12 **a** Membership functions of HR data. **b** Membership functions of GSR data

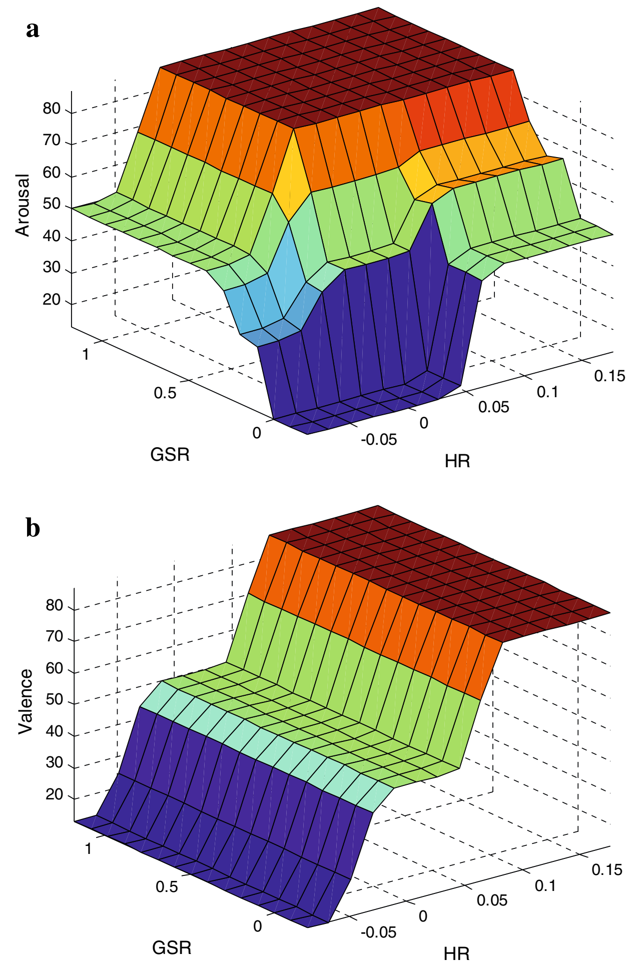


Fig. 13 **a** Response surface for arousal. **b** Response surface for valence

To validate the fuzzy model output, correlation analyses were conducted between fuzzy-inferred arousal and valence values and participant subjective ratings on the SAM questionnaire. The overall correlation between fuzzy-inferred arousal and participant arousal ratings was not significant; however, fuzzy-inferred valence significantly corresponded with participant valence ratings ($r = 0.14$, $p = 0.0144$). Results of within-feature correlations are summarized in Table 14 with the fuzzy model outputs showing a strong association with ratings of valence for the interactivity feature manipulation.

Table 14 Correlations between fuzzy inference results and subjective ratings for arousal and valence for each robot feature

Feature	Arousal correlation		Valence correlation	
	r	p	r	p
Face	−0.059	0.508	0.073	0.408
Voice	0.031	0.589	0.140	0.014
Interactivity	0.036	0.683	0.248	0.005

3.5 Correlation analysis

Correlation analyses were conducted on participant ratings of arousal and valence with the physiological measures of HR (average value for each trial) and GSR (maximum value for each trial). Results revealed a positive linear association of valence and HR ($r = 0.193$, $p = 0.0007$), but no significant correlation between arousal and GSR ($p = 0.737$). There was also a positive correlation between arousal and HR ($r = 0.141$, $p = 0.014$), but valence and GSR were not significantly correlated ($p = 0.885$).

3.6 Interview results

After the experiment trials, we asked participants for their general impressions of the use of robot technology in healthcare environments. Fifteen (62.5%) 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 did not like the robot configurations or thought they were not ready for hospital use.

When we asked for comments on the specific robot features, 15 participants (62.5%) liked the voice feature, particularly the human voice. They also pointed out that the touch-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 participants (20.8%) either did not 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.

4 Discussion

4.1 Subjective measures

In line with our expectation, the results of the between-feature analysis indicated that adding human-like features (face, voice, and interactivity) to the robot platform promoted positive emotional responses in terms of subjective measures of arousal and valence. This was in agreement with predictions from Epley et al. [57] theory that users tend to have stronger emotional experiences with a robot with anthropomorphic and interactive features. However, it was not clear which type of feature would be most effective for facilitating emotionally-rich interaction for patients. (The power function analysis provided additional information regarding the effect of features for conveying anthropomorphism, as discussed below.) A possible reason for this is that all three types of features might be important to causing emotional responses and thus participant arousal and valence for the different robot configurations was relatively consistent.

From the within-feature analyses, the effect of the various levels of robot features on PA was consistent with our hypothesis. When features became more human-like, participant ratings of PA increased. Disalvo et al. [39] and Nass et al. [44] 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 primarily due to resulting differences in user interaction caused by specific features. For example, abstract and human-like faces differed in valence ratings but not arousal. The face feature of the robot was static (no change or movement during the medicine delivery); therefore, participant perception of the face when the robot approached was the basis for the experience of valence (i.e., whether the robot was pleasurable). On the contrary, the levels of interactivity (visual message and touch-screen 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 direct commands was more exciting for participants than passively receiving the medicine without knowing what the robot would do next. The lack of significant differences among the synthesized and digitized speech conditions may have been due to condition similarities including identical message content. The length of the WAV files for both conditions was also relatively short (7–8 s) and the synthesized speech was generated by a state-of-art text-to-speech software application. Several participants commented in the final interview that they could not differentiate the two voices (synthesized and recorded human).

4.2 Physiological measures

In addition to subjective measures, the various robot features were found to affect participant physiological responses at certain trial events. For example, when the robot opened its gripper, participant HR in trials involving an interactivity feature manipulation were higher than those with a face or voice condition. This was possibly due to the excitement of participants when they realized the robot was obeying their command through the touch-screen. Within-feature analyses indicated that HR under touch-screen confirmation was higher than for visual messaging and the control condition during Events 2 (the robot stopped in front of the participant), 4 (the robot opened gripper), and 6 (the robot moved from participant sight). HR for the touch-screen confirmation and visual message conditions was also higher than the control condition in Event 5 (the robot turning and leaving the participant). These results suggest that HR might be increased by participant physical actions necessary for those events and robot configurations requiring interactivity. Other research on mental workload (e.g., [58]) also suggests that HR is higher for physical tasks than cognitive tasks.

GSR under the control condition was found to be higher than for the face condition. One possible reason for this is that in the control condition participants still saw the robot platform and the additional face features may not have had a significant additive effect on participant GSR. Another possible reason is that the GSR signal attenuates over time. It is possible that even though a participant may have experienced higher arousal under the face condition, if the condition occurred later in the experiment, the GSR response may have been attenuated compared to the control condition response.

Participant HR was also higher under the voice conditions than the face and control conditions. A possible reason might be that processing of, and responses to, auditory information is more automatic that perceiving humanoid faces, and the physiological reaction such as increase of heart rate does not depend on cognition beyond event perception. Related to this, participant HR with digitized voice messages was found to be higher than the control condition, suggesting that a familiar human voice might be more appropriate for attracting user attention.

4.3 Power function fitting

The result of the power function fitting revealed the voice feature to have the least economy for driving participant perception of robot humanness as compared to other humanoid features (i.e., face configurations). This implies making a robot sound like a real human may be less important to perceived humanness than physical appearance. However, it is also possible that our quantification of voice features in terms of frequency variance may not have reflected important characteristics of the speech capability for participants to make judgments of robot humanness. This finding merits further exploration.

4.4 Correlation analyses and fuzzy model outcomes

The experiment results showed a significant positive correlation between the subjective measure of valence and HR, but no significant correlation between arousal and GSR. There was also a positive correlation between arousal and HR, which was unexpected based on previous research results. A further correlation analysis showed a significant positive linear association of the subjective measures of arousal and valence ($r = 0.804$, $p < 0.0001$). Although it is possible that changes in HR affected both emotional responses in a similar manner, other potential explanations for the significant correlation between arousal and valence relate to experimental control. The arousal and valence rating scales were presented directly adjacent to each other on the SAM questionnaire. The order of presentation of the two scales was randomized from trial to trial; however, participants may have disregarded the scale anchors in certain trials and made marks in similar locations across scales. Furthermore, although participants were familiarized with the SAM questionnaire in the training session, there may have been differences in interpretation of the rating scales during testing.

Related to these correlations, the fuzzy modeling results showed that the subjective measure of valence was positively correlated with the valence values inferred based on the physiological measures. This provided some evidence that multiple physiological variables can be used as a basis for classifying user emotional states in a HRI context. It also provides further validation of Mandryk's [28] methods of using fuzzy logic models for classifying emotional states based on physiological measures.

As the context of the present study was HRI and the elderly participants exhibited more moderate physiological responses (as compared to younger populations), the measures we collected did not show dramatic variations for different robot features (as seen from the ANOVA results), nor did they correlate strongly with the subjective measures of arousal and valence. This may have limited the predictive power of the fuzzy models.

5 Conclusion

Considering the current increase in healthcare service demands and shortage of nurses, service robots represent a potential 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 and perceptions of the quality of healthcare services. On this basis, we investigated the effect of different robot design 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 (subjective feelings and physiological responses). Within each feature type, the general trend was that perception of humanness and positive emotional responses increased with increasing feature complexity or realism (compared to the human). Furthermore, humanoid robot features appear to differ in terms of their economy (or power) for driving participant perceived anthropomorphism and possibly emotional responses. Although these results may not be generalized to long-term interactions between users and service robots, they provide a basis for appropriate selection and implementations of robot features, including face, voice and interactivity, to promote accurate user perceptions of robot interactive or social capability and positive user first impressions.

One limitation of the present study is that there may have been carryover effects among trials due to the within-subjects experiment design. The randomization of robot conditions for each participant was aimed at reducing these carryover effects. Furthermore, the experiment design supported assessments of the variability in emotional state within a particular participant.

With respect to the subjective measures used in the study, it is possible that the two-dimensional emotion model (arousal and valence) might be relatively simple for this particular context, as evidenced by the correlation between arousal and valence measures. Future studies could measure compliance-related emotions based on appraisal theory, for example. Additional studies are also needed to further understand the relationship between robot features and user emotional experiences, in order to generate a collection of affective robot interface design guidelines for healthcare and other applications.

Acknowledgments This research was supported by the Department of Industrial and Systems Engineering at North Carolina State University. Biwen Zhu's work on the project was funded by The Ergonomics Center of North Carolina. We would like to thank the two senior centers (Glenaire and Jordan Oaks Retirement Community) in Cary, North Carolina for their help in recruiting participants and providing facilities (on-site) for our experimentation.

References

1. Auerbach DI, Buerhaus PI, Staiger DO (2007) Better late than never: workforce supply implications of later entry into nursing. Health Aff 26:178–185
2. Cleary PD, Edgman-Levitan S, Roberts M, Moloney TW, McMullen W, Walker JD, Delbanco TL (1991) Patients evaluate their hospital care: a national survey. Health Aff 10:254
3. Goodin HJ (2003) The nursing shortage in the United States of America: an integrative literature review. J Adv Nurs 43:335–350
4. Rogers AE, Hwang WT, Scott LD, Aiken LH, Dinges DF (2004) The working hours of hospital staff nurses and patient safety. Health Aff 23:202–212
5. Goetz J, Kiesler S, Powers A (2003) Matching robot appearance and behavior to tasks to improve human-robot cooperation. In: Proceedings of the 12th IEEE international workshop on robot and human interactive communication (ROMAN), pp 55–60
6. Reeves B, Nass C (1996) The media equation: how people treat computers, television, and new media like real people and places. Cambridge University Press, New York
7. Gockley R, Forlizzi J, Simmons R (2006) Interactions with a moody robot. In: Proceedings of the 1st ACM SIGCHI/SIGART conference on human–robot interaction (HRI), pp 186–193
8. Edsinger A, Kemp CC (2007) Human-robot interaction for cooperative manipulation: handing objects to one another. In: Proceedings of the 16th IEEE international symposium on robot and human interactive communication (ROMAN), pp 1167–1172
9. Choi YS, Chen T, Jain A, Anderson C, Glass JD, Kemp CC (2009) Hand it over or set it down: a user study of object delivery with an assistive mobile manipulator. In: Proceedings of the 18th IEEE international symposium on robot and human interactive communication (ROMAN)
10. Agah A, Tanie K (1997) Human interaction with a service robot: mobile-manipulator handing over an object to a human. Proc IEEE Int Conf Robot Autom 1:575–580
11. Kulic D, Croft E (2006) Estimating robot induced affective state using hidden Markov models. In: Proceedings of the 15th IEEE international symposium on robot and human interactive communication (ROMAN), pp 257–262
12. Steinfeld A, Fong T, Kaber DB, Lewis M, Scholtz J, Schultz A, Goodrich M (2006) Common metrics for human-robot interaction. In: Proceedings of the 1st ACM SIGCHI/SIGART conference on human–robot interaction (HRI), pp 33–40
13. Ekman P (1992) An argument for basic emotions. Cogn Emot 6:169–200
14. Ekman P (1992) Are there basic emotions? Psychol Rev 99:550–553
15. Lang PJ (1995) The emotion probe: studies of motivation and attention. Am Psychol 50:372–385
16. Russell JA (1980) A circumplex model of affect. J Pers Soc Psychol 39:1161–1178
17. Roseman I (2001) A model of appraisal in the emotion system. In: Scherer KR, Schorr A, Johnstone T (eds) Appraisal process in emotion: theory, methods, research. Oxford University Press, New York
18. Peter C, Herbon A (2006) Emotion representation and physiology assignments in digital systems. Interact Comp 18:139–170
19. Ekman P (1999) Basic emotions. In: Dalgleish T, Power M (eds) Handbook of cognition and emotion. Wiley, Sussex
20. Lazarus RS (2001) Relational meaning and discrete emotions. In: Scherer KR, Schorr A, Johnstone T (eds) Appraisal process in emotion: theory, methods, research. Oxford University Press, New York
21. Scherer KR (2001) Appraisal considered as a process of multilevel sequential checking. In: Scherer KR, Schorr A, Johnstone T (eds) Appraisal processes in emotion: theory, methods, research. Oxford University Press, New York pp 92–120
22. van Reekum CM, Johnstone T, Banse R, Etter A, Wehrle T, Scherer KR (2004) Psychophysiological responses to appraisal dimensions in a computer game. Cogn Emot 18:663–688
23. Russell JA, Weiss A, Mendelsohn GA (1989) affect grid: a single-item scale of pleasure and arousal. J Pers Soc Psychol 57:493–502

24. Bradley MM, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatr 25:49–59
25. Frijda NH (1986) The Emotions. Cambridge University Press, London
26. Scheirer J, Fernandez R, Klein J, Picard RW (2002) Frustrating the user on purpose: a step toward building an affective computer. Interact Comp 14:93–118
27. Ward RD, Marsden PH (2004) affective computing: problems, reactions and intentions. Interact Comp 16:707–713
28. Mandryk RL, Atkins MS (2007) A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. Int J Human Comp Stud 65:329–347
29. Guthrie SE (1997) Anthropomorphism: a definition and a theory. In: Mitchell RW, Thompson NS, Miles HL (eds) Anthropomorphism, anecdotes, and animals. State University of New York Press, Albany pp 50–58
30. DiSalvo CF, Gemperle F (2003) From seduction to fulfillment: the use of anthropomorphic form in design. In: Proceedings of the international conference on designing pleasurable products and interfaces, pp 67–72
31. Luczak H, Roetting M, Schmidt L (2003) Let's talk: anthropomorphization as means to cope with stress of interacting with technical devices. Ergonomics 46:1361–1374
32. Nass C, Moon Y (2000) Machines and mindlessness: social responses to computers. J Soc Issues 56:81–103
33. Breazeal C (2003) Emotion and sociable humanoid robots. Int J Human Comp Stud 59:119–155
34. Fong T, Nourbakhsh I, Dautenhahn K (2003) A survey of socially interactive robots. Robot Autonom Syst 42:143–166
35. Turkle S (1984) The second self: computers and the human spirit. Simon and Schuster, New York
36. Duffy BR (2003) Anthropomorphism and the social robot. Robot Autonom Syst 42:177–190
37. Bruce A, Nourbakhsh I, Simmons R (2002) The role of expressiveness and attention in human-robot interaction. In: Proceedings of the IEEE international conference on robotics and automation (ICRA), pp 4138–4142
38. Zhang T, Zhu B, Lee L, Kaber DB (2008) Service robot anthropomorphism and interface design for emotion in human-robot interaction. In: Proceedings of the IEEE international conference on automation science and engineering (CASE), pp 674–679
39. DiSalvo CF, Gemperle F, Forlizzi J, Kiesler S (2002) All robots are not created equal: the design and perception of humanoid robot heads. In: Proceedings of the 4th conference on designing interactive systems: processes, practices, methods, and techniques, pp 321–326
40. Pollack ME, Brown L, Colbry D, Orosz C, Peintner B, Ramakrishnan S, Engberg S, Matthews JT, Dunbar-Jacob J, McCarthy CE, Thrun S, Montemerlo M, Pineau J, Roy N (2002) Pearl: a mobile robotic assistant for the elderly. In: Proceedings of the AAAI workshop on automation as eldercare
41. Schaeffer C, May T (1999) Care-O-bot: a system for assisting elderly or disabled persons in home environments. In: Proceedings of AAATE-99
42. Burgard W, Cremers AB, Fox D, Hähnel D, Lakemeyer G, Schulz D, Steiner W, Thrun S (1999) Experiences with an interactive museum tour-guide robot. Artif Intell 114:3–55
43. Stern SE, Mullennix JW, Dyson C, Wilson SJ (1999) The persuasiveness of synthetic speech versus human speech. Hum Factors 41:588–595
44. Nass C, Foehr U, Brave S, Somoza M (2001) The effects of emotion of voice in synthesized and recorded speech. In: Proceedings of the AAAI symposium emotional and intelligent II: the tangled knot of social cognition, North Falmouth, MA
45. Gockley R, Matarić MJ (2006) Encouraging physical therapy compliance with a hands-off mobile robot. In: Proceedings of the 1st ACM SIGCHI/SIGART conference on human–robot interaction (HRI), pp 186–193
46. Catrambone R, Stasko J, Xiao J (2002) Anthropomorphic agents as a user interface paradigm: experimental findings and a framework for research. In: Proceedings of the 24th annual conference of the Cognitive Science Society, pp 166–171
47. Dawson ME, Schell AM, Filion DL (2007) The electrodermal system. In: Cacioppo JT, Tassinary LG, Berntson G (eds) Handbook of psychophysiology, vol 2. Cambridge University Press, London pp 159–181
48. Stevens SS (1957) On the psychophysical law. Psychol Rev 64:153–181
49. Bradley MM, Greenwald MK, Hamm AO (1993) Affective picture processing. In: Birbaumer N, Öhman A (eds) The structure of emotion. Hogrefe & Huber Publishers, Toronto pp 48–65
50. Detenber BH, Simons RF, Bennett GG Jr (1998) Roll'Em!: the effects of picture motion on emotional responses. J Broadcast Electron Media 42
51. Dautenhahn K, Walters M, Woods S, Koay KL, Nehaniv CL, Sisbot A, Alami R, Siméon T (2006) How may I serve you?: a robot companion approaching a seated person in a helping context. In: Proceedings of the 1st ACM SIGCHI/SIGART conference on human–robot interaction (HRI), pp 172–179
52. Walters ML, Dautenhahn K, Woods SN, Koay KL (2007) Robotic etiquette: results from user studies involving a fetch and carry task. In: Proceedings of the 2nd ACM SIGCHI/SIGART conference on human–robot interaction (HRI), pp 317–324
53. Lang PJ, Greenwald MK, Bradley MM, Hamm AO (1993) Looking at pictures: affective, facial, visceral, and behavioral reactions. Psychophysiology 30:261–273
54. Levenson RW (1988) Emotion and the autonomic nervous system: a prospectus for research on automonic specificity. In: Wagner HL (ed) Social psychophysiology and emotion: theory and clinical applications. Wiley, London pp 17–42
55. Ekman P (1984) Expression and the nature of emotion. In: Scherer KS, Ekman PApproaches to Emotion. Erlbaum, Hillsdale
56. Nunnally JC, Bernstein IH (1994) Psychometry theory. McGraw-Hill, New York
57. Epley N, Waytz A, Cacioppo JT (2007) On seeing human: a three-factor theory of anthropomorphism. Psychol Rev 114:864–886
58. Backs RW (1995) Going beyond heart rate: modes of autonomic control in the cardiovascular assessment of mental workload. Int J Aviat Psychol 5:25–48