Centers for Disease Control and Prevention
STEPHEN B. THACKER
CDC LIBRARY

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.

Assessing the Effects of Humanoid Robot Features on Patient Emotion during a Medicine Delivery Task

Manida Swangnetr, Biwen Zhu, Kinley B. Taylor & David B. Kaber
Edward P. Fitts Department of Industrial and Systems Engineering
North Carolina State University

Human perceptions of the "humanness" of robots have been found to be influenced by face, voice and interactivity features. These features have been studied individually in a human robot interaction (HRI) and facial features appear to be strongest in promoting positive human emotions. The objective of this study was to assess the effects of combined humanoid robot features on human emotions during a medicine delivery task. Seven robot prototypes with various combinations of face, voice and interactivity features were developed and classified in terms of levels of humanness. A "Wizard of Oz" experiment was conducted in which 32 subjects received and accepted a simulated bag of medicine from each of the robot prototypes. Both subjective (arousal and valence ratings) and physiological (HR and GSR) measures were collected as indicators of participant emotional states. Results revealed robot configurations with higher levels of humanness promoted positive emotions. Arousal and valence ratings and the HR response had utility for predicting emotions. We also found that additional humanoid features lead to higher GSR ratings, but the trend was not strictly linear with the pre-defined level of robot humanness.

INTRODUCTION

Commercial robots have been developed to assist nurses in medicine delivery services over past two decades (Krishnamurthy & Evans, 1992). Such robots normally carry medications or other healthcare related material from pharmacies to nursing stations with no direct interaction with patients. However, recent advances in intelligent control systems and precision sensors are expected to support direct robot interaction with patients in the future. For such application, there is a need to design robots with features that not only support effective task performance but also facilitate positive patient emotional experiences.

Robot anthropomorphism (i.e., the attribution of human-like qualities to non-human objects (Guthrie, 1997)) is an important concept in robot interface design for motivating user interest and perceptions of robot capabilities. Previous studies have found that anthropomorphic characteristics may elicit robot acceptance and human compliance with robot behaviors (e.g., Goetz, Kiesler & Powers, 2003), which is particularly important in patient service tasks. Based on a review of interface features implemented in existing service robots (Zhang et al., 2008), there appear to be three key anthropomorphic elements (i.e., humanoid features) to support social interaction with humans and positive human emotions, including: (1) illusion or presence of a face and/or head; (2) voice capabilities; and (3) the manner in which a user interacts with the robot (e.g., keypad to acknowledge medicine delivery or a screen displaying relevant prescription information).

In a previous study (Zhang et al., 2010), we investigated the effect of these different humanoid robot features on human perception of robot humanness and emotional responses in a medicine delivery scenario. The study evaluated the use of face, voice or interactivity features separately for conveying two levels of humanness (abstract and humanlike). Comparison was made of all robot configurations with a control condition (i.e., a robot platform without any additional features). Combinations of the three humanoid features were not prototyped and tested.

Russell's (1980) orthogonal two-dimensional space of emotion, defined by valence (pleasant / unpleasant) and arousal (high arousal / sleepiness) is a widely accepted theory of human emotion. We used this as a basis for our study and measured participant subjective reports of emotional state with the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994). In addition, heart rate (HR) and galvanic skin responses (GSR) were collected, which are two of the most commonly observed physiological responses for indicating states of arousal and valence (Dawson, Schell and Filion, 2007). Specifically, HR has been found to be positively correlated with valence (Bradley, Greenwald & Hamm, 1993) and arousal (Mandryk & Atkins, 2007). GSR has also been found to increase as the level of arousal increases (Amershi, Conati & Maclaren, 2006; Mandryk & Atkins, 2007; Dawson, Schell & Filion, 2007). Opposite to arousal, the relationship between GSR and valence is not definite (Dawson, Schell & Filion, 2007).

Our results (Zhang et al., 2010) indicated that adding anthropomorphic features to service robots promotes a greater sense of humanness and positive user emotional experiences. Within each feature type, the trend was for increased perception of humanness and positive emotional responses as robot features became more humanlike. Furthermore, humanoid robot features appear to differ in terms of their economy (or power) for driving participant perceived anthropomorphism and possibly emotional responses. Facial features were found to have the highest utility for predicting perceived robot humanness.

The objective of this follow-up study was to assess the effect of combinations of nursing robot physical appearance and interface features on perceptions of humanness and emotions in a simulated medicine delivery task. It was expected that as the robot configurations increased in their degree of humanness, this would lead to an increase in positive emotional responses, in terms of subjective valence and arousal, as well as the HR and GSR responses. Levels of humanness for feature combinations were developed based on analysis of results from prior research (see below for details).

METHODOLOGY

Thirty-two participants (19 males and 13 females) were recruited from the university population and surrounding area. Their ages ranged from 18 to 32 years (M = 23.16, SD = 3.12). None of them had any direct experience in interacting with robots or had ever seen the robot used in this study (a PeopleBot). The subjects' primary task was to receive and accept a simulated bag of medicine from the robot, interacting with it as necessary. For simplicity and control in testing, we used a "Wizard of Oz" methodology, where a researcher remotely controlled the robot's speech and movements (using a teleoperation interface).

The PeopleBot platform, by Mobile Robots Inc. (left panel in Figure 1), was used for this experiment. To simulate various levels of humanness, the basic robot platform was outfitted with different feature configurations, which included: (1) face; (2) voice; or (3) an interactive interface. Each of the three features had two levels (see right panel in Figure 1). For example, the face feature was either an "abstract face" comprised of two mini-cameras, which looked liked eyes or a "humanlike face" incorporating a smooth mask. The voice was either a pre-recorded female voice (digitized) or a computer-generated female voice (synthesized). With respect to the interactivity feature, participants were either required to read a visual message on a tablet PC mounted on the PeopleBot or they pressed a button on the touch screen to confirm medicine delivery (confirmation).

Figure 1. Basic PeopleBot platform (left). Additional anthropomorphic features (right), clockwise from top left: abstract face; human face; confirmation and; message only.

Independent Variables and Response Measures

The robot configuration was the independent variable in this study. Seven different combinations of humanoid and interactivity features were presented to participants (see Table 1). In the design of these conditions, the robot face feature was always coupled with a specific voice feature (i.e., human face with digitized voice; abstract face with synthesized voice). This condition specification was used to avoid inconsistency in robot interface design and to ensure common expectations during patient robot interaction (see for example Walters, Syrdal, Dautenhahn, Boekhorst, & Koay, 2008). Since Zhang et al. (2010) found robot facial features to have the greatest utility for predicting humanness, a prototype with a human face and digitized voice (Conditions 5, 6 and 7) was considered to represent a higher level of humanness, as compared with a robot with an abstract face and synthesized voice (Conditions 2, 3, and 4). With respect to the interactivity feature, prior results on this feature alone indicated "confirmation" to be regarded as a more humanlike behavior than conveying a "message only". The robot conditions presented in Table 1 (from 1 to 7) are considered to represent increasing degrees of humanness.

Table 1. Robot configurations defining degrees of humanness

Condition	Face	Voice	Interactivity
1	No	No	No
2	Abstract	Synthesized	No
3	Abstract	Synthesized	Visual message
4	Abstract	Synthesized	Confirmation
5	Human	Digitized	No
6	Human	Digitized	Visual message
7	Human	Digitized	Confirmation

The dependant variables included the physiological responses of HR and GSR, as well as the subjective responses of arousal and valance. Ratings of arousal and valence were collected using the SAM, which is considered to be an operational definition of Russel's (1980) emotion model (Bradley and Lang, 1994). The SAM included two questions assessing happiness and excitement levels. Participants were presented with a scale with values from 1 to 9 and picture icons associated with each value to represent emotional states. Participants identified the picture and value that best represented their emotional state during a test trial.

For the physiological data, HR was collected using a Polar S810i wristwatch receiver and transmitter chest strap. Data was collected during each trial in RR intervals. The GSR was collected using sensors attached to the index and ring fingers on the palmer side of the distal section on the non-dominant hand with a frequency of 1024 Hz.

Experiment Design and Subject Instruction

Each participant completed a total of 21 trials (7 robot conditions with 3 replications). A three factor randomized complete block design was used. The effects of subject and replication were blocked and the robot condition was regarded as a fixed effect. The conditions were assigned randomly during the experimental trials. The robot motion path and stopping area remained consistent during all trials. Figure 2 presents a diagram of the experiment facility.

Participants were provided with a training session in which the basic robot platform was used to simulate a test trial. For test trials, participants were instructed to sit still, refrain from talking, and use their dominant hand to take the medicine bag from the robot. The bag was attached to a gripper on the robot using Velcro and could only be removed when the gripper was open. They were also told to observe the different features of the robot during the interaction. They were given a stylus to use if they were required to confirm receipt of the medicine with the robot (Figure 3; Click here to view video). The HR and GSR data were collected while a participant observed and interacted with a robot. These responses were indexed to trial events (e.g., "robot enters room" using an electronic marker integrated with the data collection computer). After each trial, participants filled-out the SAM questionnaire, where they were told to think about what they felt after the robot gave them the bag of medicine.

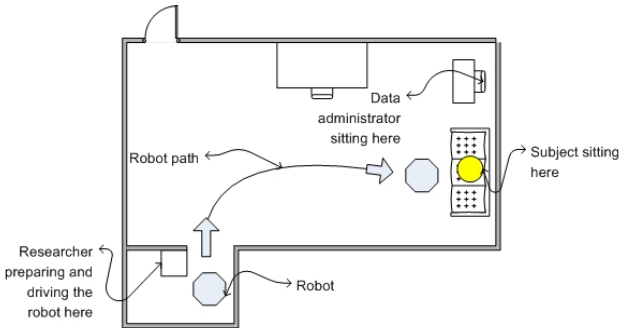

Figure 2. Diagram of experiment facility and robot path.

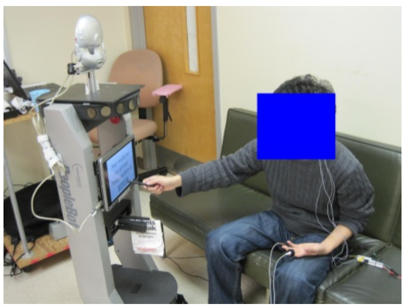

Figure 3. Subject confirming receipt of medicine from robot.

DATA ANALYSIS AND RESULTS

The subjective ratings of arousal and valence were transformed to z-scores for each subject to account for individual differences in internal scaling of emotions. Participant HR and GSR data were also normalized by calculating the changes in each response relative to the baseline or rest condition. The minimum HR value recorded during the rest period (i.e., when no stimulus was presented to subjects) was used as the baseline HR value. Similarly, we used the minimum GSR response during the rest period as the GSR baseline.

As physiological measures (HR and GSR) are generally analyzed on an event-basis, several events were identified during the robot delivery task, including the robot entering the simulated patient room; the robot stopping in front of the subject; the robot speaking to the subjects, etc. Instead of analyzing the data for all these robot events, we focused on the human responding to the robot when it opened the gripper to provide the medicine. Our previous study found that participant HR and GSR during this event provided the greatest degree of discrimination among various emotional states (Swangnetr, Kaber & Lee, 2009). This particular event is also the key event in the robot delivery service. Participants need to perceive the robot interface cues, understand how to interact with the robot, and to make a decision regarding whether to accept the medicine. This process is fairly complicated from an information processing perspective and may be more dependent on human emotions than other behaviors associated with the other scenario events.

With respect to the event time window for the GSR and HR responses, a 4-second window after the event was used based on suggestions from previous emotion research (Dawson, Schell & Filion, 2007; Lang, Greenwald, Bradley & Hamm, 1993; Levenson, 1988). 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 HR was calculated (see formulas as follows).

$$GSR_{\max} = \max_{4\text{sec.}}\left\{\frac{GSR - \min(GSR_{rest})}{\min(GSR_{rest})}\right\} \quad (1)$$

$$HR = \operatorname{mean}_{4\text{sec.}}\left\{\frac{HR - \min(HR_{rest})}{\min(HR_{rest})}\right\} \quad (2)$$

HR and GSR outliers were removed from the experiment data set before any statistical analysis. HR outliers included 10 low data points that occurred due to equipment attachment problems. With respect to the GSR data, the response variance for one subject was found to be substantially different from all others, possibly due to problems with sensor mounting or sensor errors. It is also possible that the subject experienced high anxiety during the experiment. The resulting descriptive statistics for all the responses across robot conditions are presented in Table 2.

Table 2. Descriptive statistics of response measures across robot conditions (numbers in parentheses are SD).

Cond.	HR(mean)		GSR(max)		Arousal(mean)		Valence(mean)	
1	0.230	(0.143)	0.318	(0.217)	-0.758	(0.999)	-0.727	(0.985)
2	0.251	(0.119)	0.342	(0.238)	-0.331	(0.787)	-0.283	(0.899)
3	0.255	(0.120)	0.341	(0.242)	0.136	(0.851)	0.149	(0.811)
4	0.257	(0.129)	0.369	(0.254)	-0.148	(0.851)	-0.139	(0.949)
5	0.276	(0.139)	0.336	(0.245)	0.353	(0.873)	0.434	(0.860)
6	0.310	(0.128)	0.433	(0.316)	0.170	(0.877)	0.070	(0.957)
7	0.291	(0.136)	0.410	(0.298)	0.578	(0.957)	0.497	(0.803)
Total	0.267	(0.133)	0.364	(0.262)	0.000	(0.977)	0.000	(0.977)

MANOVA Modeling

Diagnostics of the SAM ratings revealed them to be approximately normally distributed and to have equal variances across robot conditions; however, HR and GSR data did not have these characteristics. Common transformation techniques were applied to the physiological data and the normality assumption of the ANOVA was satisfied for HR (using logarithmic transformation). Therefore, a random effects MANOVA model was constructed with response measures including arousal, valence and transformed HR. GSR responses were subjected to non-parametric analysis.

Robot condition, subject and replication were included as predictors in the MANOVA model. Results showed that all three variables were significant with respect to effects on the family of response measures (Wilks' $\Lambda = 0.762$, $p < 0.0001$ for replication; $\Lambda = 0.276$, $p < 0.0001$ for subject; $\Lambda = 0.661$, $p < 0.0001$ for condition). The significant "subject" effect (after response normalization) suggested that robot stimuli had different effects on emotion across subjects. In addition, results indicated that emotion responses were time-dependent. Some subjects appeared to become bored after the second replication of test trials. However, this study did not focus on subject first impressions triggered by robot configurations with humanoid features (i.e., after they were familiarized with the PeopleBot platform). The reason for this is that patients in hospitals are likely to interact with robots repeatedly (in hallways, at nurses stations, etc.). Consequently, statistical analysis for an initial interaction effect was not conducted. Based on these findings, univariate ANOVA models were conducted to examine the main effect of robot condition on arousal, valence and HR.

ANOVA Modeling

A series of three-way (robot condition, subject and replication) ANOVAs were conducted on subjective measures of arousal, valence and HR. Results revealed that subject SAM ratings were significantly different across the seven robot conditions ($F(6, 663) = 32.06$, $p < 0.0001$ for arousal; $F(6, 663) = 24.17$, $p < 0.0001$ for valence). Post-hoc analysis using Duncan's test indicated that subject arousal and valence ratings increased as the level of robot humanness increased (see Figure 4). However, certain robot conditions could not be discriminated in terms of subject ratings (e.g., Conditions 4, 5 and 6 for arousal; Conditions 6 and 7 for valence, etc.). It is possible that subjects were less sensitive to the robot feature manipulations across these conditions. In addition, significant replication effects were also found for both arousal ($F(2, 663) = 101.41$, $p < 0.0001$) and valence ($F(2,663) = 39.15$, $p < 0.0001$), again indicating the time-dependence of the emotion responses. No significant interaction effect between replication and condition was found. There were also no significant individual differences (i.e., no subject effect), likely due to the SAM data being standardized.

ANOVA results showed that both subject ($F(31,624) = 51.47$, $p < 0.0001$) and robot condition ($F(6,624) = 20.86$, $p < 0.0001$) had a significant effect on HR. Unlike the SAM data, subject HR did not appear to be time-dependent ($F(2,622) = 2.26$, $p = 0.1057$). The post-hoc analysis results on HR across the seven robot conditions are presented in Figure 5. In general, HR increased as the level of robot humanness increased, except for Condition 4. This particular condition represented an "intermediate" level of humanness, but yielded the highest HR. This might due to the condition involving participant body movement (i.e., pointing and clicking robot interface). Such movement can accelerate HR (Backs, 1995)

Non-parametric Modeling

Friedman's two-way non-parametric test was used to assess the effect of robot condition on GSR. Subject was treated as a blocking effect in the model. As expected, the GSR response was significantly different across the seven robot conditions ($\chi^2 = 35.583$, $p < 0.0001$). However, post-hoc analysis using pair-wise comparisons revealed that GSR did not necessarily increase with increasing levels of robot humanness (see Figure 6). In general, the robot conditions requiring participants to confirm the robot medicine delivery yielded higher GSR, as compared with other conditions.

DISCUSSION

With the aging U.S. population and age-related healthcare needs, the elderly are expected to be the largest user group of service robots in the future. Although previous literature has found that older individuals are less willing to use robots than younger persons (in general), this appears to depend on individual and robot factors (Broadbent, Stafford & MacDonald, 2009). Our previous study was conducted with elderly participants, aging from 64-91 years. The post-experiment interview revealed that 62.5% of participants would like to have delivery robots working in hospitals. In addition, results from the interview in the present study indicated that more than 95% of (younger) participants would accept robots in a medicine delivery task.

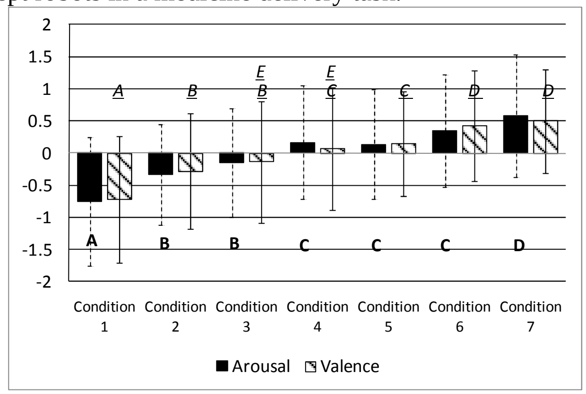

Figure 4. Post-hoc results on SAM ratings across robot conditions. (Bold letters indicate Duncan's means breakout results for arousal. Italic letters indicate results for valence. Means with the same letter are not significantly different.)

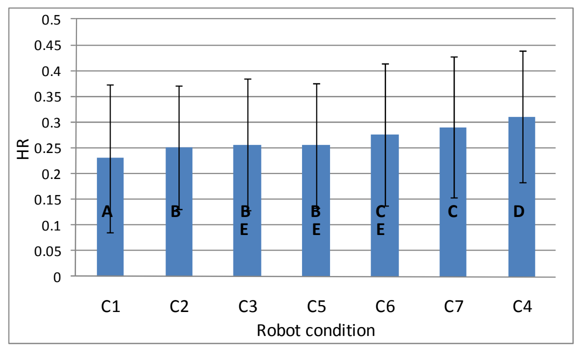

Figure 5. Post-hoc results on HR across robot conditions.

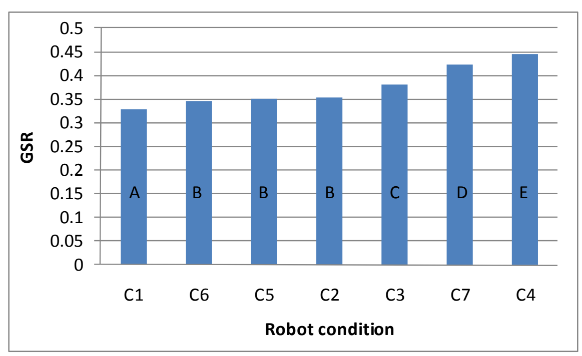

Figure 6. Post-hoc results on GSR across robot conditions.

In line with Epley, Waytz and Cacioppo's (2007) three-factor theory of anthropomorphism, our results showed that additional humanoid features in robot interface design lead to higher emotional responses in terms of SAM ratings and physiological measures. We also found that, in general, a robot with higher degrees of humanness (defined based on prior research results) would lead to higher subjective arousal and valence. An increase in both emotional responses implies higher positive emotion (Watson & Tellegen, 1985). In fact, the ordering of the mean SAM ratings across seven robot conditions had a perfect positive correlation with the pre-defined order of robot humanness (Spearman's ρ=1.0); however, the arousal ratings were swapped between Conditions 4 and 5. Post-hoc results indicated that the arousal ratings for the two conditions were actually within the same group. This means subjects may have lacked sensitivity to differences in these robot feature manipulations. Dynamic facial and vocal features should be tested in future studies.

There was also evidence from a final interview with subjects that the levels of humanness in robot design made sense to them. When defining the levels of humanness, we prioritized face and voice feature manipulations and considered interactivity feature manipulations as ancillary in triggering human emotion responses. Such arrangement was consistent with the participants' post-test report regarding their opinions of the robots. Twenty-six out of 32 felt that face and voice were the most important to facilitate positive emotional responses during HRI.

Participant bodily movements during trials (i.e., confirming robot delivery of the medicine) lead to the highest physiological responses in Conditions 4 and 7. It was also interesting to note that HR and GSR in Condition 4 were significantly higher than those in Condition 7. We suspected that although "clicking a button" added a certain amount of interactivity during the HRI, this form of interactivity was more indicative of human-machine interaction vs. human-human interaction. Therefore, subjects were less comfortable when the human face and digitized voice were coupled with "clicking and confirming" using the robot interface. In future studies, more intuitive interactivity (e.g., a robot opening a gripper when a patient reaches for medication) should be manipulated and tested.

No evidence was found to indicate that GSR increased with increasing levels of robot humanness. Conditions 5 and 6 failed to yield higher GSR responses even though they were associated with higher levels of humanness. This might have been due to the participants feeling less excited (low arousal and low GSR) when the robot with a human face and digitized voice had a very limited way of interacting with them (either no interaction or presentation of a visual message). The approach to interaction may not have met participant expectation for a service delivery robot and, therefore, lead to lower GSR responses.

Besides a possible physical exertion effect from interactivity features (i.e., Conditions 4 and 7), the HR response increased as the degree of robot humanness increased. Although there was no evidence indicating that GSR responses increased as the level of robot humanness increased, the amplitude of the GSR signal was higher when participants interacted with a robot with additional anthropomorphic features than with a robot without. Therefore, changes in the HR and GSR responses reflected an increase due to positive emotions when participants interacted with a humanoid robot.

CONCLUSION

In conclusion, the current study provides an initial basis for the design of service robots integrating combinations of humanoid features for a medicine delivery task. Our findings support the notion that additional anthropomorphic features in robot design trigger positive human emotional responses in terms of both subjective and physiological measures. This may in turn lead to increased patient acceptance of robots and compliance with robot requests. Our design of robot prototypes, and categorization in terms of levels of humanness, were successful and had some utility for predicting human arousal, valence and HR responses. Future studies should focus on testing a broader sample population including actual patients in healthcare facilities. Dynamic facial features and more intuitive interactivity features should also be explored in service robot configuration.

REFERENCES

Amershi, S., Conati, C., & Maclaren, H. (2006). Using feature selection and unsupervised clustering to identify affective expressions in educational games. In *Proc Workshop Motivational and Affective Issues in ITS, 8th Int Conf on Intell Tutoring Syst*, (pp. 21-28). Jhongli, Taiwan.

Backs, R. W. (1995). Going beyond heart rate: Modes of autonomic control in the cardiovascular assessment of mental workload. *The Int J Aviat Psychol*, 5, 25-48.

Bradley, M. M., Greenwald, M., K., & Hamm, A. O. (1993). Affective picture processing. In N. Birbaumer, & A. Öhman, *The Structure of Emotion* (pp. 48-65). Toronto: Hogrefe & Huber Publishers.

Bradley, M., & Lang, P. (1994). Measuring emotion: the self-assessment Manikin and the semantic differential. *J Behav Ther Exp Psychiatr*, 25 (1), 49–59.

Broadbent, E., Stafford, R. & MacDonald, B. (2009). Acceptance of healthcare robots for the older population: review and future directions. *Int J Soc Robotics, 1*, 319–330.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The Electrodermal System. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson, *Handbook of Psychophysiology* (3rd edn., pp. 159-181). NY: Cambridge University Press.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychol Rev*, 114(4), 864-886.

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation, In *Proc 12th IEEE Int Workshop Robot and Human Interact Commun*, (pp. 55-60).

Guthrie, S. E. (1997). Anthropomorphism: A definition and a theory. In R. W. Mitchell, N. S. Thompson, & H. L. Miles, *Anthropomorphism, Anecdotes, and Animals* (pp. 50-58). Albany: SUNY Press.

Krishnamurthy, B., & Evans, J. (1992). HelpMate: a robotic courier for hospital use. *IEEE Int Conf on Syst, Man and Cybern*, Chicago, IL.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3), 261-273.

Levenson, R. W. (1988). Emotion and the autonomic nervous system: A prospectus for research on automonic specificity. In H. L. Wagner (Ed.), *Social Psychophysiology and Emotion: Theory and Clinical Applications* (pp. 17-42). London: Wiley.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *Int J Hum Comput Stud*, 65, 329–347.

Russell, J. A. (1980). A circumplex model of affect. *J Pers Soc Psychol*, 39 (6), 1161-1178.

Swangnetr, M., Kaber, D. & Lee, Y-S. (2009). Human emotional state classification based on wavelet analysis and statistical feature selection. In *Proc IEA2009 17th World Congress on Ergonomics*. Beijing, China.

Walters, M. L., Syrdal, D. S., Dautenhahn, K., Boekhorst, R. T., & Koay, K. L. (2008). Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Autonomous Robots*, 24 (2), 159-178.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychol Bull, 98*, 219-235.

Zhang, T., Kaber, D., Zhu, B., Swangnetr, M., Mosaly, P. & Hodge, L. (2010). Service robot design effects on user perceptions and emotional responses. *J Intell Serv Robotics*, 3 (2), 73-88.

Zhang, T., Zhu, B., Lee, L., & Kaber, D. B. (2008). Service robot anthropomorphism and interface design for emotion in human-robot interaction, In *Proc IEEE Int Conf Autom Sci Eng*, (pp. 674-679).