16th IEEE International Conference on Robot & Human Interactive Communication
August 26 ~ 29, 2007 / Jeju, Korea
WC2-3

Investigating Implicit Cues for User State Estimation in Human-Robot Interaction Using Physiological Measurements

Emily Mower † ‡ David J. Feil-Seifer † Maja J. Matarić † Shrikanth Narayanan ‡

† *Interaction Laboratory*
Department of Computer Science
University of Southern California
Los Angeles, CA USA 90089-0781
mower|dfseifer|mataric@usc.edu

‡ *Speech Analysis and Interpretation Laboratory*
Department of Electrical Engineering
University of Southern California
Los Angeles, CA USA 90089-0781
shri@sipi.usc.edu

***Abstract*— Achieving and maintaining user engagement is a key goal of human-robot interaction. This paper presents a method for determining user engagement state from physiological data (including galvanic skin response and skin temperature). In the reported study, physiological data were measured while participants played a wire puzzle game moderated by either a simulated or embodied robot, both with varying personalities. The resulting physiological data were segmented and classified based on position within trial using the K-Nearest Neighbors algorithm. We found it was possible to estimate the user's engagement state for trials of variable length with an accuracy of 84.73%. In future experiments, this ability would allow assistive robot moderators to estimate the user's likelihood of ending an interaction at any given point during the interaction. This knowledge could then be used to adapt the behavior of the robot in an attempt to re-engage the user.**

I. INTRODUCTION

A robot can fulfill various roles in the course of an interaction with a human. One of the roles of most interest in our work is that of coach [5]. As a coach, a robot interacts with a user in order to supervise and guide that user to improve performance on a task or skill. For instance, the robot may interact with an individual recovering from stroke, spending extended time with the patient in order to improve the pace and scope of recovery.

Socially Assistive Robotics (SAR) is the intersection of Socially Interactive Robotics, or robotics where interaction with humans is through social interaction, and assistive robotics, or robotics whose primary purpose is assistance [5]. SAR focuses on robotics for assistive applications that achieves that assistance through social, rather than physical, interaction. One of the fundamental goals of SAR is to create stimulating and engaging environments and interactions in which a user willingly participates for an extended period of time. This is inherently difficult since every user has specific interactive needs. It is therefore necessary for an interactive agent to recognize the needs of the user and adapt its behavior to meet those needs. However, it cannot be assumed that the user will always choose to, or even be able to, express those needs. This lack of explicit communication has led many researchers to investigate the utility of implicit user state cues, such as physiological responses [12].

In the past, robotic coaches have relied on movement cues [4] to determine user state. However, physical user state is not perfectly correlated with emotional state. As research in physiological data analysis progresses, the ability to infer such uncommunicated user state through the use of physiological data cues will continue to improve. It is with this goal in mind that the presented experiment was designed.

This experiment is an exploratory analysis of the utility of physiological data processing for user state identification. We were interested in how physiological signals could be used to estimate various aspects of user state, including engagement. If the robot were able to estimate engagement in terms of both the user's performance and the user's implicit physiological cues, it could be endowed with the ability to identify state change such as a change indicating that the user would be terminating an exercise. The robot could then appropriately alter its interaction in an attempt to return the user to an engaged state and prevent the termination of the exercise.

Our experimental design was based on a therapy exercise currently used for individuals recovering from a stroke. In this rehabilitation exercise, participants are presented with a contorted wire puzzle and a metal loop, and instructed to move the loop along the wire puzzle without touching the loop to the puzzle (Figure 1(a)). We used the wire puzzle exercise to produce variations in user emotion state and user performance. There were a total of two simple moderator types (a simulated and an embodied robot) each with three personality types (positive, negative, and neutral). The moderators had simple physical gestures and appearance to allow us to form a baseline understanding of the correlation between user performance, user engagement/boredom, user physiological expression, and moderator personality in the course of a human-machine interaction. The physiological data signals chosen include galvanic skin response (GSR) and skin temperature.

This paper will show that physiological data, specifically GSR and skin temperature, can be used to predict user engagement state. We found that it was possible to identify a single user's engagement state with an average accuracy of 83.87% using the engagement state data from the cohort population.

978-1-4244-1635-6/07/$25.00

II. Related Work

A. Physiological Data and Emotion

The physiological expression of emotion is controlled by the neural circuitry of the brain. Emotions influence the majority of physical function and are influenced by both the state of the body and stimuli applied to the body [17]. Emotional states are defined such that, for a specific state, there exists a probable set of somatic and autonomic nervous system outputs.

Electrodermal activity (EDA), commonly measured using Galvanic Skin Response (GSR), is the result of the secretion of sweat by the eccrine gland. The primary function of this gland is thermoregulation, but it has shown to be more responsive to significant, including emotional, stimuli than to thermal stimuli. As the level of sweat in the gland changes, the conductivity and resistivity properties of the skin change as well. The higher the sweat rises in the gland, the lower the skin resistance measured, leading to differing EDA measurements. The EDA measurements are controlled almost entirely by the sympathetic nervous system (a component of the autonomic nervous system) [1].

Skin temperature fluctuations are also associated with autonomic system activity. These fluctuations are caused by changes in the flow of blood resulting from arterial blood pressure or vascular resistance. Like GSR, skin temperature has been found to be an effective method of emotional state estimation [8].

B. Emotion Classification

One of the difficulties in creating a classifier based on emotional states is the ambiguity surrounding the definitions of such states. Emotion classification remains a difficult problem due to the non-linear mapping between human emotional state and human expression (facial affect, body language, vocal prosody, etc.), as well as the presence of purposeful deception and cultural emotional differences [2]. This occurs even in the presence of predefined, highly specified categorical labels.

When presented with neutral-content speech, humans are able to classify the emotion state of the speaker with approximately 60% accuracy [12]. When these participants are presented with facial expressions, they are able to classify the affective state between 70-80% of the time. Although computers have been found to classify with much higher rates [12], the rates may be further improved by including physiological data that are more difficult for the user to consciously manipulate.

Many studies analyzing the physiological bases of emotional responses have utilized sensory presentation methods to elicit emotional responses [8], [11], [12]. In these studies, a subject is presented with a combination of images, audio, and lighting conditions to evoke specific emotional responses. Using heart rate, galvanic skin response (GSR), and temperature, Lisetti et al. were able to successfully classify frustration (out of six possible emotion classes) with 77.3% accuracy [11]. However, it is not clear that this type of experimental design produces emotional responses similar in nature to natural expressions of emotion. To address this problem, several studies have utilized games designed to frustrate the user [3], [14]. Scheirer et al. [14] used skin conductivity, blood volume pressure, and a pressure mouse to determine the frustration state of the users within a game that was designed to crash randomly. The group found that frustration could be accurately predicted 67.4% of the time in the particular experimental context. Since an individual's level of engagement is based on many emotional (in addition to task) factors, we hypothesized that the findings from the emotional classification studies could be extended to this related domain. This extension has been used previously [13], [15]. Rani et al. compared the performance of various machine learning techniques using 46 features derived from cardiac activity, heart sound, bioimpedance, electrodermal activity, electromyographic activity, and skin temperature to identify states of anxiety, engagement, boredom, frustration, and anger in real time. They were able to correctly classify (across all emotion states) 85.81% of the time using Support Vector Machines [13]. The work described in this paper differs from Rani et al. in two important ways. Firstly, the task described in this paper was not simulated but involved physical three-dimensional manipulations. Secondly, in this paper the environment in which the users participated was inherently interactive while the experiments described in Rani et al. were primarily task-focused.

III. Hypotheses

The goal of this experiment was to determine if user emotion state, specifically level of engagement, could be predicted using physiological signals. A user's engagement state (his willingness to continue) is important information pertaining to an interaction. Physiological data have shown to be a good predictor of user interest/frustrative state. We therefore hypothesized that since there exist interpersonal similarities with respect to physiological state, physiological data from a set of users would be a good predictor of the engagement state of a user from the same population.

H1: It is possible to determine the user's engagement state using the engagement state data from the other users in the cohort population.

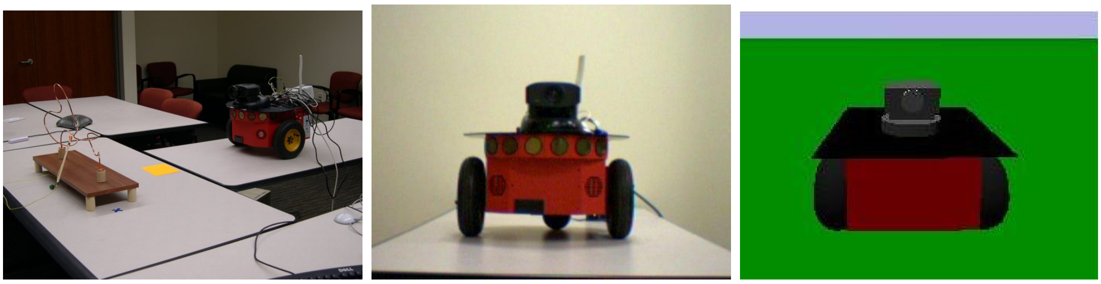
One must also expect that the performance of the user will depend on the current behavior of the robot. In therapeutic and experimental settings, the performance of specific users has been shown to be dependent on the moderator personality [10], [16]. It has also been shown that task compliance is affected by the robot's personality and that participants are more compliant with a serious robot than with a more playful robot [7]. In this case, performance success/task compliance is defined by the number of faults accrued (where the goal is to accrue the fewest faults). We hypothesized that a robot's outward personality would have an effect on task performance.

H2: Individuals interacting with a negative (more serious) moderator would have the best performance when compared to individuals interacting with the neutral or positive (more playful) moderator.

IV. Wire Puzzle Game

A. Experimental Design

1) Game Description: This experiment was based on a wire puzzle game (Figure 1(a)). There were three trials each separated by a minute-long break. In a given experiment,



(a) Wire puzzle and robot.

(b) The ActivMedia Pioneer 2 DX robot used in the experiments. The motions of the PTZ camera provide feedback to the user that supplements the audio.

(c) The standard Gazebo simulation of the embodied robot.

Fig. 1. The experimental conditions.

each of the trials used the same moderator type (simulated or embodied robot) but a different personality type (positive, negative, neutral).

The participants were told that the goal of the game was to move a wire loop from one end of the puzzle to the other without touching the loop to the wire puzzle structure while accumulating the lowest score possible in a given subtrial. The score was dependent both on the subtrial completion time and the accumulated number of faults. A subtrial was completed every time the wire was moved from one end of the puzzle to the other. Every time the wire loop touched the puzzle, the participants received a fault. A fault caused the participant's score to be incremented dependent upon the amount of contact time. A trial was defined as the completion of as many subtrials as the participant desired. The participants could end a trial at any point by pushing a button. Therefore, the engagement state of the user could be defined by the time-on-task. The first third of the trial was defined as the most engaged while the final third (the time leading up to the point when the user decided to end the trial) was defined as the least engaged.

2) Moderator Interaction: The purpose of the moderator (from a functional perspective) was to serve as a real-time performance feedback device. At the beginning of each trial, the moderator informed the user of the rules of the game. Within the course of a subtrial, the moderator indicated each time the user received a fault. Upon completion of a subtrial, the robot informed the user of his time and position on the high score list.

The users were presented with a different personality style in each trial (positive, negative, and neutral). The moderators were designed to induce a satisfied, frustrated, and neutral emotion state for a user. The motion of the robot was kept to a minimum to avoid a confounding effect on the emotional state of the user.

There were two moderator types: an embodied robotic moderator (shown in Figure 1(b)) and a virtual computer moderator. The computer moderator was a simulation of the robotic moderator (shown in Figure 1(c)). The vocal prompts, physical gestures, and sensing capabilities of the two moderators were identical and used the same software. The users interacted with only one of the two moderator types.

The moderator behavior was broken down into three categories: positive, negative, and neutral. The behavior cases were differentiated based on vocal behavior, physical movement, and feedback. All of the vocal prompts were recorded by a professional actress. The positive behavior case used a bright tone of voice and provided strong encouragement. The negative behavior case used a scornful tone of voice and provided little encouragement. The neutral behavior case used a neutral tone of voice and provided basic encouragement. The positive behavior case moved its camera up and down while vocalizing (to express approval), the negative behavior case moved its camera from side to side while vocalizing (to express disapproval), and the neutral behavior case did not move its camera. Each participant was presented with all three personality types in randomized order.

The moderator behavior was also differentiated based on fault notification. The notification for a negative fault was a navy whistle, for a neutral fault it was a simple ding, and for a positive fault it was a short clip of water bubbling. In the negative behavior case, the moderator would indicate a fault if no significant event (fault or puzzle completion) had occurred within the last four seconds. In the positive behavior case, the moderator would neglect to include and notify the user of one-third of the faults accrued. In the neutral behavior case, the moderator accurately reported all faults accrued.

The initial high score table was identical for each user. In the negative case it was weighted such that it was very difficult for the user to achieve the top slots, while in the positive case, the table was weighted such that the user could more easily reach the top position.

3) Physical Platform: The wire puzzle was composed of a copper wire contorted into a path (Figure 1(a)). In order to determine contact or lack thereof, all inputs from the puzzle were connected through a USB port to the robot. The ends of the puzzle were marked by wooden dowels with copper coils.

We used an ActivMedia Pioneer 2DX mobile robot equipped with a Sony PTZ camera (see Figure 1(b)). The robot was controlled through Player [6], an abstraction layer allowing the same code to be used on both the robot and the simulator moderator conditions.

The simulated robot was rendered using Gazebo [9], which creates a 3D rendering of its worlds and dynamics.

Gazebo contains an approximate physical model of the Pioneer robot and the PTZ camera used in the robot condition of the experiment (Figure 1(c)). The simulator is controlled through Player.

B. Participant Selection

Participants were drawn from the students, faculty, and staff at USC. There were a total of 26 subjects (6 female, 20 male). They were recruited via email and word of mouth. No financial compensation was offered. The participants were assigned to either the computer moderated or the robot moderated experiment such that there was equal representation across moderators. Six of the subjects were rejected due to device failure (2 female, 4 male).

Fig. 2. BodyMedia SenseWear Pro2 Armband for measurement of GSR and skin temperature

C. Data Acquisition

Four types of data were recorded from each participant: audio, visual, physiological, and performance. The audio and visual data were recorded using a video camera. The physiological data were recorded using the BodyMedia ®PRO_2 Armband (Figure 2). The participants wore the armband on the upper arm such that the armband was in contact with their skin throughout the duration of the experiment. The BodyMedia device collects GSR and skin temperature at a rate of 4 samples per second. The performance data consisted of statistics derived from the trials, including the number of subtrials per trial, number of faults per subtrial, and time per subtrial.

V. Results and Analysis

The analysis used the physiological and task performance data, as the speech data were too sparse to allow for any significant results. Hypothesis one (H1) was investigated using the physiological data and hypothesis two (H2) was evaluated using the performance data.

A. Exploration of Hypothesis One

H1: It is possible to determine the user's engagement state using the engagement state data from the other users in the cohort population.

The feature vector used in the statistical analysis of the physiological data consisted of the GSR mean, minimum value, and maximum value, and skin temperature mean, variance, maximum value, and range over three-second (12-sample) windows [11], [14]. It was necessary to normalize the data in order to compare across individuals. To do so, each of the ten statistics was altered such that each statistic over a single user had a mean of 0 and range between -1 and 1. This allowed us to compare variations across users, since each individual's physiological response range is different. We then employed the K-Nearest Neighbors (KNN) algorithm on the physiological data feature set.

In the analysis, each of the three trials (positive, negative, and neutral personality) was segmented into thirds based on the user-dependent length of each of the specific trials. Since each of the three trials were of user-dependent length, it was possible to describe the resultant length based on engagement state. The first third of each trial was defined as the highest engaged state as the user was furthest (temporally) from quitting. The second third of the trial was defined as moderately engaged, and the final third was defined as the least engaged because the user quit while in this third of the trial. This hypothesis would be supported if it was possible to estimate, given a 3-second segment of data, from which third (first, second, or third) the segment of data originated.

The first step in this analysis was to ensure that the trends observed in the GSR and skin temperature data when the user was participating in the trial were significantly different than the trends observed when the subject was not participating in a trial (Figure 3). The between-trial data were acquired in the time between the trials when the participant sat quietly for one minute. This comparison served to indicate whether the statistical trends seen in the trial data were artifacts or were reflective of user engagement state.

To test this, we formed six groups of statistics:

- Group one represented an aggregate of 3-second segments from the first third of trials one, two and three.
- Group two was an aggregate of 3-second segments from the second third of trials one, two, and three.
- Group three was an aggregate of 3-second segments from the final third of trials one, two and three.
- Group four was an aggregate of the first third of the between trial data.
- Group five was an aggregate of the second third segments of the between trial data.
- Group six was an aggregate of the final third segments of the between trial data.

The features were analyzed using a two-way analysis of variance (ANOVA) comparing similar thirds (i.e., the first third column is a comparison between groups one and four, see Table I for the skin temperature analysis). The resulting analysis indicates that the trends observed are significant. For example, in the first row of Table I the skin temperature mean is shown to be significantly different in the between and within trial cases for the first third (significant, $F[1, 5490] = 32.075, p < 0.001$), second third (significant, $F[1, 5490] = 14.669, p < 0.001$), and final third (significant, $F[1, 5490] = 2.876, p < 0.001$). A similar result can be seen graphically in the raw GSR data (Figure 3). This indicates that the trends measured in the within trial settings were neither device artifacts nor a natural resting human tendency but were instead due to participation in the task.

The data were then classified using K-Nearest Neighbors and a feature vector composed of the statistics discussed above to analyze the hypothesis. A single subject's data (from a single trial) were used as test data and the remaining

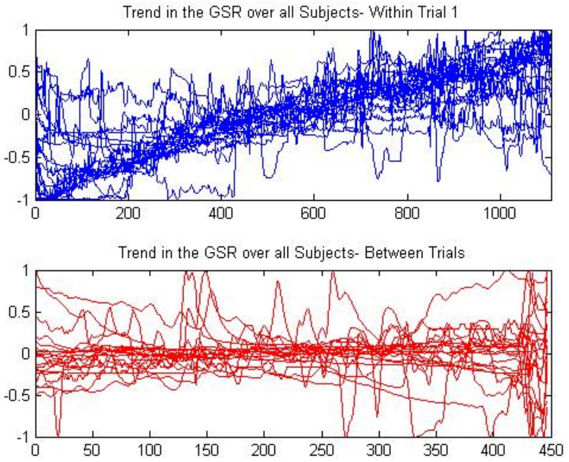


Fig. 3. Top plot: GSR trend of all subjects in trial 1. Bottom plot: GSR trend of subjects between trials 1 and 2.

Stat	1^{st} Third $F[1,5490]$	2^{nd} Third $F[1,5490]$	3^{rd} Third $F[1,5490]$
Mean	32.075*	14.669*	27.876*
Var	1236.148*	841.223*	916.329*
Max	103.322*	18.795*	0.040
Range	1499.003*	1412.874*	1356.126*

TABLE I

ANOVA ANALYSIS COMPARING TEMPORAL GROUPINGS OF SKIN TEMPERATURE STATISTICS ($n = 20$): GROUP 1 VS. GROUP 4 (1^{st} THIRD), GROUP 2 VS. GROUP 5 (2^{nd} THIRD), AND GROUP 3 VS. GROUP 6 (3^{rd} THIRD). THE STATISTICS MARKED WITH A $*$ WERE FOUND TO BE SIGNIFICANT ($p < 0.001$).

subjects' (19 in total) data (again from the same trial and across both embodiment types) were used to train the KNN ($k = 15$) algorithm. The results were validated using leave-one-out cross-validation. The results from each trial were averaged. Across all the subjects, it was possible to estimate the third of the trial from which the data came with an accuracy dependent on trial number. It was possible to determine the position: within trial one with an accuracy of 84.73% (see Table II for the confusion matrix); within trial two with an accuracy of 80.94%; and within trial three with an accuracy of 77.05%.

The next logical step was to determine whether the engagement state of a single subject's 3-second window of data from a specific trial could be predicted using the data from the remaining two trials. If this prediction were possible, it would indicate that the robot was capable of learning the engagement state behavior of a specific user. The answer to this question was determined in a similar way to the previous question. Again the trial data were tagged by position (first, second, or final third) and the two unqueried trials were used to train the KNN algorithm. We found that, across all subjects, it was possible to estimate the position within trial one given data from trials two and three with an average accuracy of 76.47%; trial two given data from trials one and three with an average accuracy of 80.57%; trial three given data from trials one and two with an average accuracy of 79.94%.

	1^{st} Third	2^{nd} Third	3^{rd} Third
1^{st} Third	0.8577	0.0682	0.0741
2^{nd} Third	0.0989	0.8246	0.0744
3^{rd} Third	0.0719	0.0687	0.8594

TABLE II

CONFUSION MATRIX FOR TRIAL ONE: PARTICIPANT INDEPENDENT ENGAGEMENT TEST, CLASSIFIED (VERTICAL) VS. ACTUAL (HORIZONTAL)

B. *Exploration of Hypothesis Two*

H2: Individuals will accrue fewer faults when interacting with a negative moderator than when interacting with a non-negative (positive or neutral) moderator.

There were a total of 20 subjects. The average trial length (independent of trial number and moderator type) was 365.14 seconds. The average number of subtrials was 9.23 for the robotic trial and 9.41 for the computer trial. The average number of subtrials for trial one was 10.2, for trial two was 8.25, and for trial three was 8.55. The average number of subtrials for the positive moderator was 8.3, for the negative moderator was 9.15, and for the neutral moderator was 9.55.

The second hypothesis was tested using a one-way ANOVA. The average number of fault instances (per subtrial) when interacting with the negative moderator was 17.96 the number of fault instances with the positive moderator was 27.28 and with the neutral moderator was 28.58. The number of fault instances with the negative moderator was lower than the number of instances with the positive moderator (significant at $\alpha = 0.05, F[1,38] = 4.176, p = 0.048$) and with the neutral moderator (significant at $\alpha = 0.05, F[1,38] = 5.168, p = .029$). The difference between the number of faults accrued in the positively and neutrally moderated case was not significant ($F[1,38] = 0.051, p = 0.822$).

VI. DISCUSSION AND CONCLUSIONS

The main finding of this work is the position prediction framework. Within our framework, it was possible to identify the position information (first, second, or final third) within the context of a trial, with high precision. The ability to classify the position within a trial (an estimation of engagement state) is important in human-robot interaction because it allows a robot to estimate whether or not a subject is about to quit an interaction. This ability could be highly valuable in human-robot interaction (HRI). In the rehabilitation context, the robot could monitor an individual's progress in an exercise and use the information about the user's intention to quit to appropriately adapt its behavior and motivation strategy in order to attempt to reengage the user. This type of dynamic analysis of behavior will also allow researchers to analyze the success of their engagement solutions.

In this rehabilitative context, task performance has been used to motivate the exercises. The accuracy of these predictions may be greatly improved when physiological data are utilized. Without an indication of implicit state, it may be difficult for the robotic moderator to determine if continued efforts with no discernable improvement are the indication that the user is about to quit or if instead it is an indication that the user is committing the skills to memory. When the physiological data are added to the model, the robot will be able to determine if the signal profile is more indicative of interest or boredom.

The models in this experiment were formed based on two subject population types, single subject (the testing and training were both with respect to a single user) and multiple subject (the testing was based on a single subject, the training on the other 19 subjects). We found that the classification rate based on single subject engagement data was slightly lower than the multiple-subject classification. The models for the engagement state classification in the single subject scenario were formed based on only two trials while in the multiple subject case they were formed based on 19 trials. Furthermore the models in the single-subject classification task did not take into account trial number. The classifier based on many users for the same trial (where each user had previously participated in the same number of trials) produced more accurate results. In the future, this classification result could be improved. If the engagement state data could be trained with a longer term interaction over a variety of tasks, the models could be better informed with respect to task-independent user behavior. The results could also be improved by incorporating task performance into the engagement state data. Therefore, if the user had completed trials of decreasing performance and the physiological data pointed to a state of lowered engagement, the robot could infer that the user would soon be quitting.

We were able to support that, in a negatively moderated case, a user's performance (in terms of lack of faults accrued) will improve. Future work will establish a further connection between robot personality and performance. For example, what happens to these effects when HRI occurs over a longer interval of time? How is this improvement affected by setting (i.e. in the lab vs. in the home)? Or, how does personality affect a user's desire to continue working with the robot?

VII. Acknowledgments

We extend our thanks to Joy Nash for lending her voice to the robot and Carlos Busso for his insight. This work was supported by the USC Institute for Creative Technologies and the NSF Graduate Research Fellowship Program.

References

[1] J. T. Cacioppo and L. G. E. Tassinary. *Principles of Psychophysiology: Physical, Social, and Inferential Elements*. Cambridge University Press, 1990.

[2] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, pages 32–80, January 2001.

[3] A. R. Damasio. *Descartes' Error: Emotion, Reason and the Human Brain*. Grosset/Putnam, 1994.

[4] J. Eriksson, M. Matarić, and C. Winstein. Hands-off assistive robotics for post-stroke arm rehabilitation. In *Proc. of the International Conference on Rehabilitation Robotics*, pages 21– 24, Chicago, Il, Jun-Jul 2005.

[5] D. Feil-Seifer and M. Matarić. Defining socially assistive robotics. In *Proc. of the International Conference on Rehabilitation Robotics*, pages 465–468, Chicago, Il, Jul 2005.

[6] B. Gerkey, R. Vaughan, and A. Howard. The player/stage project: Tools for multi-robot distributed sensor systems. In *Proceedings of the International Conference on Advanced Robotics*, pages 317–323, Coimbra, Portugal, Jul 2003.

[7] S. Kiesler and J. Goetz. Mental models and cooperation with robotic assistants. In *Proc. of Conference on Human Factors in Computing Systems*, pages 576–577, Minneapolis, Minnesota, USA, April 2002.

[8] K. H. Kim, S. W. Bang, and S. R. Kim. Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, 2004.

[9] N. Koenig and A. Howard. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2149–2154, Sendai, Japan, Sep 2004.

[10] K. M. Lee and C. Nass. Designing social presence of social actors in human computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, volume 5, pages 289 – 296, Ft. Lauderdale, Fl, April 2003.

[11] C. L. Lisetti and F. Nasoz. Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP Journal on Applied Signal Processing*, pages 1672–1687, September, 1 2004.

[12] R. W. Picard, E. Vyzas, and J. Healy. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001.

[13] P. Rani, C. Liu, and N. Sarkar. An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Analysis & Applications*, 9(1):58–69, May 2006.

[14] J. Scheirer, R. Fernandez, J. Klein, and R. W. Picard. Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers*, 14(2):93–118, February 2002.

[15] S. Scotti, M. Mauri, S. Cerutti, L. Mainardi, and M. Villamira. Quantitative evaluation of distant student psychophysical responses during the e-learning processes. In *Proceedings of International Engineering in Medicine and Biology Society*, pages 1196–1199, Shanghai, China, September 2005.

[16] A. Tapus and M. Matarić. User personality matching with hands-off robot for post-stroke rehabilitation therapy. In *Proceedings, International Symposium on Experimental Robotics (ISER)*, Rio de Janeiro, Brazil, Jul 2006.

[17] F. Toates. *Biological Psychology: An Integrative Approach*. Pearson Education Limited, Essex, England, 2001.