Survey of Psychophysiology Measurements Applied to Human-Robot Interaction

Cindy L. Bethel[1], Student Member, IEEE, Kristen Salomon[2], Robin R. Murphy[3], Member, IEEE and Jennifer L. Burke[4]

[1,3,4]Institute for Safety Security Rescue Technology, University of South Florida, Tampa, FL, USA,
e-mail: (clbethel, murphy, jlburke4)@cse.usf.edu
[2] Department of Psychology, University of South Florida, Tampa, FL, USA, e-mail:salomon@cas.usf.edu

***Abstract*—This paper reviews the literature related to the use of psychophysiology measures in human-robot interaction (HRI) studies in an effort to address the fundamental question of appropriate metrics and methodologies for evaluating HRI research, especially affect. It identifies four main methods of evaluation in HRI studies: (1) self-report measures, (2) behavioral measures, (3) psychophysiology measures, and (4) task performance. However, the paper also shows that using only one of these measures for evaluation is insufficient to provide a complete evaluation and interpretation of the interactions between a robot and the human with which it is interacting. In addition, the paper describes exemplar HRI studies which use psychophysiological measures; these implementations fall into three categories: detection and/or identification of specific emotions of participants from physiological signals, evaluation of participants' responses to a robot through physiological signals, and development and implementation of real-time control and modification of robot behaviors using physiological signals. Two open research questions on psychophysiological metrics were identified as a result of this review.**

I. Introduction

Human-robot interaction is an emerging field of research; however the development of methods to evaluate the effectiveness of these interactions is lacking. In the early phases of this field the focus was on the development of the specific robotic systems and applications. Methods of testing and evaluation have been adopted and modified from such fields as human-computer interaction, psychology, and social sciences [1]. The manner in which a human interacts with a robot is similar but not identical to interactions between a human and a computer or a human interacting with another human. As robots become more prevalent in day-to-day life, it will be increasingly important to have accurate methods of evaluating how humans feel about their interactions with robots and how they interpret the actions of the robots.

There are four main methods of evaluation used for human-robot interaction (HRI) studies: (1) *self-report measures*, (2) *behavioral measures,* (3) *psychophysiological measures,* and *(4) task performance* [1, 2]. The most common methods utilized in most HRI studies are self-report and behavioral measures. There is limited research in the use of psychophysiological measures and task performance metrics in HRI studies. Each method has its advantages and disadvantages; however some of the disadvantages can be overcome by using more than one method of evaluation [1].

The use of participants' self-reports is one of the most commonly used methods of evaluation in HRI studies and often included as part of a psychophysiological evaluation. Self-report measures include pencil-and-paper or computerized psychometric scales, questionnaires, and/or surveys. Participants provide a personal report of their motives and feelings about an object, situation, or interactions. Self-reports provide valuable information but there are problems with validity and corroboration. Participants may not answer exactly how they are feeling but rather answer questions as they feel others would answer them or in a way they think the researcher wants them to answer. Another issue with self-reporting measures is the inability for observers to corroborate immediately and directly the information provided by participants [3]. Participants may not be in touch with what they are feeling about the object or situation and therefore may not report their true feelings. The responses could be dependent on participants' mood and state of mind on the day of the study [3, 4]. For these reasons, it is important to perform psychophysiological measures to add another dimension of understanding of participants' responses and physiological reactions in HRI studies.

Behavioral measures are probably the second most common method of evaluation in human-robot interaction studies and often included in psychophysiological evaluations for convergent validity of participants' self-report responses and measured physiological reactions. Johnson and Christensen [4] define observation as "the watching of behavioral patterns of people in certain situations to obtain information about the phenomenon of interest." The "Hawthorne effect" is an area of concern with observations. If participants know that they are being observed, it will impact their behaviors [3, 4]. For this reason, psychophysiological measures can assist with obtaining a better understanding of the participants' underlying responses expressed at the time of the observations. The benefit of using behavioral measures is that researchers are able to record the actual behaviors of participants and do not have to rely on participants to report accurately their intended behaviors or preferences in addition to obtaining psychophysiological measures for convergent validity [3, 5].

The design of a quality research study for use in HRI applications that produces results that are verifiable, reliable, and reproducible is a major challenge. Psychophysiological measurements can complicate this process because the results are not always straightforward and confounds can lead to misinterpretation of data. There is a tendency to attribute more meaning to results because of the tangible nature of the recordings. Information needs to be obtained from participants prior to beginning a study to help reduce these confounds (e.g., health information, state of mind). Multiple physiological signals should be used in order to find correlations in the results. Steinfeld et al. [5] describe the need for the development of common metrics as an open research issue in HRI. They discuss an approach of developing common metrics for HRI; however this approach is oriented more toward an engineering perspective and does not completely address the social interaction perspective. Both perspectives have value but require further investigation. In order to obtain credibility in the research community, HRI studies need to be supported by quality experimental designs with adequate sample sizes and multi-faceted methods of measurement to provide convergent validity.

None of these measures alone are sufficient to interpret accurately the responses of participants to a robot with which they are interacting. In order for a study to have corroboration and consistency in its evaluations, at least two methods of measurement should be used [1, 3, 4]. Steinfeld *et al.* [5] discuss an approach of developing common metrics for human-robot interaction; however this approach is geared more towards an engineering perspective and does not completely address the social interaction perspective and evaluation. Both perspectives have value but more in-depth investigation is required.

This review begins with some general information and terminology related to psychophysiological measures (Section II). In Section III, coverage is given to three categories of human-robot interaction implementations that have utilized psychophysiological measures. A discussion of two open research questions discovered as a result of this review is presented in Section IV.

II. General Psychophysiology

Psychophysiology focuses on the interaction between the mind and body [6]. John Stern defined psychophysiology as "any research in which the dependent variable (the subject's response) is a physiological measure and the independent variable (the factor manipulated by the experimenter) a behavioral one" [6]. The most common measures in use in Human-Robot Interaction studies include: cardiovascular system *(heart rate variability (HRV), respiratory sinus arrhythmia (RSA), cardiac output, interbeat interval (IBI), blood pressure (BP))*; electrodermal activity *(skin conductance activity (SCA), skin conductance response (SCR));* respiratory system *(breaths per minute, respiration volume)*; muscular system *(electromyography (EMG));* and brain activity *(electroencephalography (EEG) and imaging)* [6-9].

There are two types of psychophysiological response tendencies, (1) *individual-response stereotypy* and (2) *stimulus-response specificity* that commonly occur in psychophysiological studies; however they are not mutually exclusive. *Individual-response stereotypy* occurs when a few individuals exhibit a pattern of responses different than expected to a specific stimulus or stressor. Also, individuals may have the same idiosyncratic response to different stressors, no matter what the stressors may be. *Stimulus-response specificity* is when a stimulus or stressor produces a similar pattern of physiological responses among most subjects or participants studied. Typically, more than one type of response is involved but the pattern of responses would be consistent among most participants subjected to the same stimulus or stressor [6].

In most psychophysiological studies, there are three primary responses that are measured: (1) *tonic response,* (2) *phasic response,* and (3) *spontaneous/non-specific response.* The *tonic responses* are the baseline or resting level responses of activity for a particular physiological measure. This is a level that occurs when participants being measured are not making responses to a known or unknown stimulus. The *phasic response* occurs when participants have discrete responses to a specific or known stimulus (*an evoked response*). It is important during this type of measurement to account for internal in addition to external stimulus that may impact participants' responses to the presented stimuli. This can be accomplished through self-reports or interviews to make sure other factors (e.g., participants' state of mind or mood) are not contributing to the measured responses. A *spontaneous/non-specific response* is a measurable response when there is no known stimulus presented [6].

The following three response factors that need to considered in any psychophysiological study: (1) *orienting response,* (2) *defensive response,* and (3) *startle response*. The *orienting response* relates to how a participant responds to novel stimuli. It causes the participant to orient towards the novel stimuli to identify what it is and its location. Once the participant determines this is not a threat or some concerning stimuli, the effects of the orienting response are inhibited. Therefore, the first few seconds of the presentation of novel stimuli should in some cases be disregarded when evaluating the collected data depending on the application. There are some cases where researchers may want to evaluate or measure the *orienting response* toward a robot presented to participants. The *defensive response* occurs as a result of intense, threatening, dangerous, or painful stimulus. This type of response prepares the body for "fight or flight" activation in a participant. The inclusion of this data would depend on the type of study conducted. The *startle response* occurs due to a sudden onset of an intense type of stimuli (e.g., door slam or lightning strike). Data collected after a *startle response* would be handled similar to an *orienting response* by disregarding the data for the first few seconds following the presentation of the stimuli; however this would be dependent on the focus of the research study [6].

Habituation reduces participants' responses due to repetitive presentation of the same or similar stimulus in psychophysiological studies. There are two types of habituation: (1) *short-term* – occurs during a single evaluation session and (2) *long-term* – occurs over multiple settings over a period of days or weeks. *Habituation* occurs more rapidly when a stimulus is presented frequently. One method to slow down the process of *habituation* is to ask participants to complete a rating questionnaire between the presentations of each stimulus to initiate a behavioral response. *Habituation* has its strongest effects towards the end of any study and needs to be considered in the evaluation of any data collected during a psychophysiological study [6].

There are advantages of using psychophysiological measures in human-robot interaction applications and experiments. The primary advantage is that participants cannot consciously manipulate the activities of their autonomic nervous system [1, 10-14]. Additionally, psychophysiological measures offer a non-invasive method that can be used to determine the stress levels and reactions of participants interacting with technology [10-14].

The use of psychophysiological measures can pose significant challenges. The ability to gather reliable data from participants in real-world human-robot interaction scenarios can be difficult [1]. Ambulatory systems have been developed that will accurately record physiological data while participants are mobile, although adjustments must be made for movement artifacts in the data. Proper preparation of the area where electrodes are placed, location of electrode placement, and making sure appropriate amounts of conducting gel or paste are used are factors which impact the quality of data collected.

It is important and sometimes complicated to determine baseline values; and the law of initial values can make this issue even more problematic [6, 8-10]. The "Law of Initial Values" indicates that the initial state of a physiological system determines the level of possible changes in that state that can occur [6]. If the system is measured at a higher initial state, then it limits further increases in physiological response levels, similarly if a system starts at a lower initial state, it will limit further decreases in levels for that system.

Controlling for confounds that could make interpretation of signals difficult to attribute to particular states or emotions is another issue that must be considered when utilizing psychophysiological measures [1, 6-10]. A common problem with psychophysiological measures is that because the outputs are tangible signals, those interpreting these results have a tendency to infer meaning that may not be accurate.

III. Implementations of Psychophysiological Measures

Implementations using psychophysiological measurements in HRI fall into three primary categories: (A) participant emotion detection and/or identification based on physiological measures [10, 15-18], (B) evaluation of participants' physiological responses to technology [13, 19], and (C) real-time robot control and behavior modifications based on physiological responses from participants [11, 12, 14]. There is limited research related to the use of psychophysiological measures in the robotics community. Of the ten implementations discussed in this review, only five involve psychophysiological measurements of participants in direct interaction with some type of robot [11-14, 19].

A. Implementations using psychophysiological measures for emotion detection and/or identification

There are five primary research studies that focus on using psychophysiological measurements to detect and/or identify specific emotions expressed by participants. In most cases, these studies are preliminary investigations to form the basis of further human-robot interaction studies and/or the development of a control architecture or behavior system.

Picard, Vyzas, and Healey [10] focus their study on the development of a machine that can accurately recognize eight distinct human emotions given four physiological signals. They discuss that machine intelligence must also include emotional intelligence. This study investigates a common issue in multiple session psychophysiological measurements which is the problem of "day-to-day" variations in a participant's emotional responses. For accurate affect recognition they feel it is important to include multiple types of signals from the participant, and to obtain information related to the participant's context, situation, goals, and preferences [10, 20]. They used a single-participant multiple-day data collection method. The participant that was used was an actress who expressed eight different emotions: no emotion (neutral), anger, hate, grief, platonic love, romantic love, joy, and reverence. The actress not only expressed each emotion externally, but focused on feeling each emotion internally. The experiments included 25-minute daily sessions across 20 days. The five physiological signals recorded were electromyography (EMG) of the masseter facial muscles, blood volume pressure (BVP), heart rate (HR), skin conductance (SCR), and respiration. The physiological signals were processed using Sequential Floating Forward Search (SFFS), Fisher Projection (FP), and their own Hybrid SFFS with Fisher Projection (SFFS-FP). The results using the SFFS-FP algorithms indicated that they obtained an 81% recognition accuracy for the eight categories of emotions which is higher than machine recognition of affect from speech (60-70%) and almost as accurate as automated recognition of affect from facial expressions (80-98%) [10, 20]. These results were significant because findings discussed in the psychophysiology literature indicated that only arousal levels could be detected through the use of psychophysiological measures [10, 20].

Rani *et al.* [15] focused their initial study on the idea that if a robot can detect stress quickly, then it can respond to the human in real-time. A robot was not used in their initial study but instead they had participants play video games and assessed their stress levels through self-report, heart rate variability and interbeat interval (IBI). A frequency domain analysis was performed of the IBI signal to detect whether participants were experiencing stress. They analyzed the results of participants' stress and developed a robotic architecture to control a robot based on psychophysiological inputs received from the human with which the robot is interacting. The developed robot architecture included a collection of electrocardiography (ECG) signals from participants and the calculation of the IBI. Next, they performed a wavelet transformation. The authors then calculated standard deviations of the sympathetic and parasympathetic frequency bands. These standard deviations became input variables into a fuzzy logic system and the output was a stress index value. If the stress index value was greater than a threshold value, the robot would receive an alarm signal and would take action to assist participants. Some problems encountered were simulating stressful situations that elicited the appropriate response, day variability in participants, and variability between participants [15].

Sarkar [18] proposed an approach and performed some initial experiments for a system that would enable a robot to recognize the psychological state of a human with which it is interacting and modify its actions or behaviors to make the human more comfortable interacting with the robot. This study was based on the assumptions that the affective state of the participant was directly related to the interaction with the robot and that only psychophysiological measures were used to recognize affect. These assumptions are limiting and not realistic but necessary for the tractability of the study. A goal of this study was to recognize human affect through the use of psychophysiological measures. The next goal was to identify the robotic actions associated with the measured affective state and modify the robot's actions to alter the affective state of the human with which it was interacting. The final goal was to design control rules for the robot to associate actions with the resulting affect expressed by the human with which it was interacting. The study utilized HRV, EMG of the cervical trapezius muscle, temperature analysis, SCR, and ECG signals. They created an online stress detection algorithm that was based on ECG signals and the power spectrum of the IBI derived from the ECG signal to obtain frequency bands for the sympathetic and parasympathetic activity of the ANS. This data was processed using fuzzy logic to form the basis of the control architecture described in [15].

Rani, Sarkar, Smith, and Adams in [17] focus their continued research studies on affect recognition based on physiological measures obtained from a wearable biofeedback sensor system. The study included six participants and was a fully within subjects design. The participants were given two versions of three problem solving tasks (solving anagrams, math problem solving, and sound discrimination) of varying difficulty across six experimental sessions to induce participant anxiety. They measured ECG, SCR, EMG of the corrugator supercilii (left brow) and masseter (jaw) muscles, skin temperature, and relative pulse volume. Self-reports were also utilized to corroborate physiological data collected with participant anxiety levels reported. The physiological signals were processed using fuzzy logic along with decision tree learning for affect detection. The data was divided into two sets, one for training the system and the other for testing the system. The results indicate they were able to detect anxiety reliably in participants involved in the problem solving sessions. They found that the decision tree learning classification system was more reliable than the fuzzy logic system of classification.

Kulić and Croft [16] began their series of research studies by estimating participant intent using physiological signals and performed some preliminary tests. They felt that by determining participants' intent through physiological measures, the robot could gain a better understanding of participants' rating of its performance without having to poll participants repeatedly for explicit feedback. They discussed the importance of using more than one physiological signal for determining participant intent accurately. For the purpose of this study they used a valence/arousal system of evaluating intent. They measured blood volume pressure, SCR, chest cavity expansion/contraction, and EMG of the corrugator supercilii (eyebrow) muscle. They processed the signals using a fuzzy inference engine with five sets of rules. The first set of rules evaluates the relationship between SCR and arousal. The second set of rules looks at the relationship between EMG and valence. The third set of rules correlates cardiac activity to valence and arousal. The fourth set of rules relates vasomotor responses to arousal. The fifth set of rules correlates respiratory activity with emotional state [16]. The experiments in this study used a picture-based system and followed the psychophysiological testing and measurement procedures developed by Lang *et al.* [21]. The procedures consisted of a baseline measurement taken for participants and then they were shown an emotionally arousing image for ten seconds and then were asked to rate the emotional content of the image using valence and arousal scales [16, 21]. The study was performed using four participants and in one case the EMG electrodes were not properly attached and the data was not usable. On average, arousal was correctly detected 94% of the time. The change of valence was correctly detected on average 80% of the time. When the valence was correctly detected, 75% of the time the direction of the valence was correctly detected. They used the results of this study to develop a robot planning and control strategy which would interact with a human and respond to the human's emotional arousal in real-time. They used the collected data to train their fuzzy logic planning and control system. They found that respiration rates were not useful in determining participants' arousal responses because the response time is too long for real-time application. Additionally, they determined that changes in heart rate were difficult to associate with a specific event or context [16]. Skin conductance response showed a linear correlation to arousal and was shown to be an effective measure. Results also indicated a relationship between the EMG measures of the corrugator supercilii muscle with valance in participants.

B. Implementations using psychophysiological measures for evaluation of participant reactions to technology

There are two primary studies that utilize psychophysiological measures to evaluate how participants respond to robotic implementations and behaviors [13, 19]. Both studies were conducted by Kulić and Croft to determine how participants would react to their robotic manipulator arm.

The two studies performed by Kulić and Croft [13, 19] utilized a robot manipulator arm and evaluated participants for their anxiety levels while experiencing various movements of the robotic arm. The robot performed two sets of movements: (1) pick and place, and (2) reach and retract. There were also two scenarios for each movement type; a set of classic potential fields planned motions and a set of safe planned motions. There were two goals associated with the first study, (1) participants' subjective and physiological responses to the robot motions, and (2) determine if a particular set of robot motions could reduce participants' anxiety levels [19]. The goals of the second study were to validate a previously developed inference engine [16] with a statistically significant sample size (36 participants); to develop and test a reliable system for determining the participants' responses to the robot motions; and whether the perception of safe motions related to the type of motion path planning used [13]. In both studies they measured heart rate, SCR, and EMG of the corrugator supercilii (eyebrow) muscle. The authors determined that participants' arousal responses could be most reliably detected with SCR, but heart rate had a contributory impact, although less reliable. Psychophysiological responses were compared with participants' self-reports. EMG of the corrugator supercilii muscle was not a reliable predictor of participants' valence (positive or negative) and arousal level in the interactions between the robot and the participant. In most participants no changes were noted. The results indicated that participants had lower arousal responses with the safe planned motions of the robotic manipulator arm and felt calmer when the robot motions were slower. Participants tended to show strong, measurable physiological responses to fast robotic arm movements. The results also indicated that physiological signals provided useful information and added a level of perceived safety for humans interacting with robots.

C. Implementations using psychophysiological measures for real-time robot control and behavior modifications

Three primary studies have been conducted related to the use of psychophysiological measures for the development and implementation of real-time robot control architectures and adaptation of robot behaviors [11, 12, 14].

A study by Rani *et al.* [11] involved the development of a robotic system that monitored a participant's anxiety level and would respond appropriately to assist the participant. They used a subsumption architecture in which the robot would normally operate in the wandering mode; however if the robot received an affect (e.g., high anxiety level) signal from the participant it would stop the wandering behavior and either rush to the aid of the participant or ask the participant questions depending on the level of affect signal detected. If the robot encountered an obstacle or something that threatened its survival it would cease all other behaviors to attend to its survival then return first to any affect signals detected and then to a wandering mode. The participant played video games of differing difficulty to induce different affect levels. They used self-report questionnaires and performed measurements for heart rate variability (HRV), IBI, skin conductance response (SCR), and electromyography (EMG) of the corrugator supercilii (eyebrow) and masseter (jaw) muscles. The study results indicated that cardiac activity, SCR, and EMG were all good indicators of anxiety and correlated with the participant's self-report. One limitation of the study was that only one participant was used in the experiments conducted.

Itoh *et al.* [12], developed their own bioinstrumentation system to measure human stress when interacting with a fixed humanoid robot that had only an upper body. Their wearable system measured ECG, respiration, EDA (changes in skin resistance), pulse wave transit time, blood pressure, and upper body movements. The experiments relied heavily on IBI derived from ECG to measure the activity of the sympathetic (LF-HRV) and parasympathetic (HF-HRV or RSA) divisions of the ANS. If participants' stress level increased past a certain threshold then the robot would modify its actions to decrease participants' stress levels by shaking the participants' hand. The physiological responses indicated a reduction in participants' stress after the robot shook their hand. Their system would modify the robot's behaviors in real-time in response the physiological data collected from participants. Results indicated that blood pressure and pulse wave transit time were disrupted due to movement artifacts and the data was not useful; however the ratio of sympathetic/parasympathetic activity was valuable in detecting participants' stress levels during their interactions with the robot [12].

Lui *et al.* [14] performed a study in which a robot modified its behavior based on the psychophysiological responses of the person with which it was interacting. In this study 14 participants performed two different versions of robot-based basketball (RBB), counterbalanced. In one version, the game difficulty was based on participants' performance and in the other version the game difficulty was based on participants' psychophysiological readings for anxiety. As participants' anxiety level increased the difficulty level would decrease and vice versa. The modification of the game occurred in real-time in response to participants' anxiety levels obtained from psychophysiological data collected. The study used self-report questionnaires of anxiety in addition to measuring cardiovascular activity (IBI, relative pulse volume, pulse transit time, and pre-ejection period), SCR (tonic and phasic), and EMG activity (from the corrugator supercilii (eyebrow), zygomaticus (corner of the mouth), and upper trapezius (shoulder) muscles). The results indicated that 11 out of 14 participants had lower anxiety levels playing the psychophysiological-based version of RBB that adjusted difficulty by participants' measured anxiety levels. Additionally, nine participants of 14 had improved performance scores with the psychophysiological-based version of RBB [14].

IV. Discussion

Two open research questions were discovered as a result of this review of the current literature related to the use of psychophysiological measures in human-robot interaction: (1) *what are the most appropriate psychophysiological signals to use in human-robot interaction studies?* and (2) *how can we verify the applicability and accuracy of psychophysiological measures in human-robot interaction studies?* There has been limited research addressing both questions, but there does not appear to be a consensus in the findings; therefore further investigation is needed.

Regarding open question one, the research indicates that it is important to utilize more than one psychophysiological signal in studies in order to have corroboration of the data and for more accurate detection and interpretation of participants' responses. The studies mentioned in this review typically have utilized more than one psychophysiological signal and there is some discussion in these studies of which signals worked better than others; though there does not appear to be a consensus as to which signals are most appropriate for HRI studies. It may be a matter of the application and purpose of the interaction, but further investigation should be conducted. Additionally, convergent validity with other types of measures such as self-reports and observations would be desirable in human-robot interaction studies.

Related to open question two, it is evident from this review that only one of the studies utilized a statistically significant number of participants (Kulić and Croft [13]) with 36 participants to provide a level of experimental reliability and validity. In some of the studies the number of participants was not given, and of the ones that were given, the number of participants ranged from one to fourteen. When the numbers of participants are so few, it is difficult to verify the accuracy and the meaningfulness of the data and results presented. Further investigations should be conducted with appropriate power analyses to determine proper sample sizes for future psychophysiological studies in human-robot interaction.

V. Conclusions

Although only limited research has been conducted in the use of psychophysiological measures in human-robot interaction, from the literature it appears that psychophysiology is a valuable tool that can be useful in human-robot interaction studies. The limited research indicates value in using psychophysiological measures as a method of evaluating a human's response to the robot with which it is interacting, in addition to the ability to use these measures to control and adjust the manner in which a robot interacts with a human in real-time. These studies have shown that psychophysiological measures can be useful in detecting and identifying specific emotions of a human that is interacting with a robot.

The literature on the use of psychophysiology in human-robot interaction studies indicates that only limited research has been devoted to the use of this system of measurement. Only five of the ten studies presented in this review actually utilized a robot interacting directly with a human [11-14, 19]. Of the studies presented, only the Kulić and Croft study included a significant number of participants with 36 individuals [13]. In the other studies presented, if the number of participants was given, the range was from one to 14 which makes it difficult to validate the reliability of the results presented.

If combined with other forms of measurement for convergent validity, it appears that psychophysiological measures can be a valuable evaluation tool for the robotics community. Steinfeld *et al.* [5] have identified the lack of common metrics in human-robot interaction as an open issue. One method of evaluation and measurement is not going to be sufficient for a complete evaluation of a human's response to a robot in human-robot interaction studies [1]. Instead, research should focus on developing a diverse set of complimentary measures that capture the full range of human-robot interactions.

VI. ACKNOWLEDGEMENTS

This work is supported in part under a National Science Foundation Graduate Research Fellowship – Award Number DGE – 0135733. Additionally, this work was sponsored, in part, by the US Army Research Laboratory under Cooperative Agreement W911NF-06-2-0041. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ARL or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

VII. REFERENCES

[1] C. D. Kidd and C. Breazeal, "Human-Robot Interaction Experiments: Lessons Learned," in *Proceeding of AISB'05 Symposium Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction*, Hatfield, Hertfordshire, UK, 2005, pp. 141 - 142.

[2] J. L. Burke, R. R. Murphy, D. R. Riddle, and T. Fincannon, "Task Performance Metrics in Human-Robot Interaction: Taking a Systems Approach," in *Performance Metrics for Intelligent Systems*, Gaithersburg, MD, 2004.

[3] D. G. Elmes, B. H. Kantowitz, and H. L. Roediger III, *Research Methods in Psychology*, Eighth ed. Belmont, CA: Thomson-Wadsworth, 2006.

[4] B. Johnson and L. Christensen, *Educational Research Quantitative, Qualitative, and Mixed Approaches*, 2nd ed. Boston, MA: Pearson Education, Inc., 2004.

[5] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common Metrics for Human-Robot Interaction," in *1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, Salt Lake City, Utah, USA, 2006.

[6] R. M. Stern, W. J. Ray, and K. S. Quigley, *Psychophysiological Recording*, 2nd ed. New York, NY, USA: Oxford University Press, 2001.

[7] M. E. Dawson, A. M. Schell, and D. L. Filion, "The Electrodermal System," in *Handbook of Psychophysiology*, 2nd ed, J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, Eds. USA: Cambridge University Press, 2000, pp. 200 - 223.

[8] K. A. Brownley, B. E. Hurwitz, and N. Schneiderman, "Cardiovascular Psychophysiology," in *Handbook of Psychophysiology*, 2nd ed, J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, Eds. USA: Cambridge University Press, 2000, pp. 224 - 264.

[9] L. G. Tassinary and J. T. Cacioppo, "The Skeletomotor System Surface Electromyography," in *Handbook of Psychophysiology*, 2nd ed, J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, Eds. USA: Cambridge University Press, 2000.

[10] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 23, pp. 1175 - 1191, 2001.

[11] P. Rani, N. Sarkar, C. A. Smith, and L. D. Kirby, "Anxiety detecting robotic system - towards implicit human-robot collaboration," *Robotica,* vol. 22, pp. 85 - 95, 2004.

[12] K. Itoh, H. Miwa, Y. Nukariya, M. Zecca, H. Takanobu, S. Roccella, M. C. Carrozza, P. Dario, and T. Atsuo, "Development of a Bioinstrumentation System in the Interaction between a Human and a Robot," in *International Conference of Intelligent Robots and Systems*, Beijing, China, 2006, pp. 2620 - 2625.

[13] D. Kulić and E. Croft, "Physiological and subjective responses to articulated robot motion," *Robotica,* vol. Forthcoming, p. 15, Aug. 08, 2006.

[14] C. Liu, P. Rani, and N. Sarkar, "Affective State Recognition and Adaptation in Human-Robot Interaction: A Design Approach," in *International Conference on Intelligent Robots and Systems*, Beijing, China, 2006, pp. 3099 - 3106.

[15] P. Rani, J. Sims, R. Brackin, and N. Sarkar, "Online stress detection using psychophysiological signals for implicit human-robot cooperation," *Robotica,* vol. 20, pp. 672 - 685, Nov. 21, 2002.

[16] D. Kulić and E. A. Croft, "Estimating Intent for Human-Robot Interaction," in *The 11th International Conference on Advanced Robotics (ICAR2003)*, Coimbra, Portugal, 2003, pp. 810 - 815.

[17] P. Rani, N. Sarkar, C. A. Smith, and J. A. Adams, "Affective communication for implicit human-machine interaction," in *IEEE International Conference on Systems, Man, and Cybernetics*, 2003, pp. 4896 - 4903.

[18] N. Sarkar, "Psychophysiological control architecture for human-robot coordination-concepts and initial experiments," in *IEEE International Conference on Robotics and Automation (ICRA'02)*, 2002, pp. 3719-3724 vol.4.

[19] D. Kulić and E. Croft, "Anxiety detection during human-robot interaction," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*, 2005, pp. 616 - 621.

[20] T. W. Bickmore and R. W. Picard, "Towards Caring Machines," in *Conference on Human Factors in Computing Systems CHI '04 extended abstracts on Human factors in computing systems*, Vienna, Austria, 2004, pp. 1489-1492.

[21] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "Emotion, Arousal, Valence, and the Startle Reflex," in *The structure of emotion : psychophysiological, cognitive, and clinical aspects*, N. Birbaumer and A. ·Ohman, Eds. Seattle: Hogrefe & Huber, 1993, pp. vi, 312.