Operator Engagement Detection for Robot Behavior Adaptation

Pramila Rani and Nilanjan Sarkar
Research & Development, Metis Design Corporation, Cambridge MA 02138
Mechanical Engineering, Vanderbilt University, Nashville TN 37235
pramila71@yahoo.com

Abstract: *It has been shown that in human-robot interaction, the effectiveness of a robot varies inversely with the operator engagement in the task. Given the importance of maintaining optimal task engagement when working with a robot, it would be immensely useful to have a robotic system that can detect the level of operator engagement and modify its behavior if required. This paper presents a framework for human-robot interaction that allows inference of operator's engagement level through the analysis of his/her physiological signals, and adaptation of robot behavior as a function of the operator's engagement level. Peripheral physiological signals were measured through wearable biofeedback sensors and a control architecture inspired by Riley's original information-flow model was developed to implement such human-robot interaction. The results from affect-elicitation tasks for human participants showed that it was possible to detect engagement through physiological sensing in real-time. An open-loop teleoperation-based robotic experiment was also conducted where the recorded physiological signals were transmitted to the robot in real-time speed to demonstrate that the presented control architecture allowed the robot to adapt its behavior based on operator engagement level.*
Keywords: *operator engagement, situation awareness, physiological sensing, human-robot interaction*

1. Introduction

Robotic technology has made commendable progress in recent years, which has ushered in many new areas of application (e.g., battlefield, space, personal assistance etc.). However, one of the major stumbling blocks in deploying completely autonomous robots in these complex and unstructured task domains is that the current robots are not fully reliable and smart enough to do such complex jobs without any human help. For instance, Project Alpha, a U.S. Joint Forces Command rapid idea analysis group, suggests that it may not be before the year 2025 that robots would be capable of completely replacing humans on the battlefield (Schafer, R., 2003). Thus, there is a real need in foreseeable future to synergistically combine various capabilities of robotic systems with the human's intelligence and cognitive task understanding so that together they can address many of the current goals of various applications.

When a human works with a robot (either as a peer or as a supervisor) the role of human error becomes significant towards the performance of the task. Human error is an important contributing factor in some of the most disastrous accidents in history where humans and machines work together (Reason, J., 1990). For example, a detailed analysis of road safety (Treat, J. R. et al, 1977) found that human error was the sole cause in 57% of all accidents and was a contributing factor in over 90% of these cases. Surprisingly, only 2.4% of these accidents were due exclusively to mechanical fault and only 4.7% of them were caused only by environmental factors. Many other studies have reported similar results. Humans, due to the inherent limitation of their information processing ability make mistakes; therefore, it is not shocking to know that human error has also been implicated in a variety of day-to-day occupational accidents, including 70% to 80% of those in civil and military aviation (O'Hare, D. et al, 1994)0. There is no systematic study found in the literature that investigated human error in the context of human-robot operations. However, studies in other disciplines such as aviation where human and machine work together have found that over the past 40 years the number of aviation accidents that were exclusively due to mechanical failure has decreased significantly, even though the decline of human error related accidents has been noticeably slower (Shapell, S. & Wiegmann, D., 1996)

It has been stressed in a number of studies including those that involved human-robot interactions (Drury, J. L. et al, 2003) that maintaining situation awareness is the key to reducing human errors. Situation awareness can be defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future (Endsley, M. R., 1988) ." There are three main components that constitute situation awareness. These are: perception, engagement and memory. While perception and memory are more difficult to control and manipulate, it may be possible to capture and retain the attention or engagement of a person in a given task for a given time.

In the context of human-robot interaction, it is important to monitor the attentional state of the human by the robot in order to detect the level of engagement as well as take meaningful actions when necessary. There could be many situations in a human-robot mission where the human as a supervisor issues a command and observes a primarily autonomous execution of the task by the robot. He/she intervenes only in exceptional cases (e.g., when the robot fails to resolve the situation by itself). All other times he/she remains as an observer. Such a sequence of short activity followed by a long inactivity on the part of the operator may cause lack of attention or engagement that may be detrimental. We believe that if robots could detect engagement/attention of the operator and alert him/her when there is a loss of engagement, we would be a step closer towards realizing truly interactive, human-like robots.

In this paper, we present our work on measuring operator task engagement though physiological feedback and integrating engagement-sensing capability in a robotic architecture. We present a teleoperation based robotic experiment in which the robot behavior adapts to the change in operator attention/engagement using recorded physiological signals. Teleoperation was chosen as an example to demonstrate such a capability because it is a classic example of "Human as controller" problems, where the lack of engagement on the part of the operator can be detrimental to the robot and the task at hand. (Fong, T. W., 2001) While significant work has been done to deal with the problems of communication delay, intermittency, and efficiency of human–robot interaction, to our knowledge our work is the first of its kind that targets behaviors adaptation of a teleoperated robot based on operator engagement level. In the context of this paper, the terms engagement and attention have been used synonymously since the term engagement essentially means "employment of attention" (www.biology-online.org).

The paper is organized as follows: Section II describes some selected research done in the past in the field of psychophysiology and human-robot interaction in teleoperation tasks. Section III describes the objective of the paper and gives details of the problem statement. Section IV describes various aspects of engagement detection – cognitive task design, measurement and processing of physiological signals, and use of regression trees for engagement recognition. Section V discusses engagement-based teleoperation - the control architecture for the present work and teleoperation-based task design. The experimental results are given in Section VI. Finally, Section VII summarizes the contribution of the paper and provides conclusions and future work plans.

2. Related Research

In this section we present a brief overview of the work done in two major areas that encompass the scope of this paper, namely, research in psychophysiology aimed at detecting mental state of an operator, and the human robot interaction related research in teleoperation.

There is a rich history in the human factors and psychophysiology literature to understand occupational stress, operator workload (Kramer, A. F., 1987)0, operator, mental effort and other similar mental states based on physiological measures such as those derived from electromyography (EMG), electroencephalography (EEG), and heart rate variability (HRV). Multiple psychophysiological measures such as HRV, EEG, blink rates and others have been used together in recent years to assess pilots' and drivers' workload (Wilson, G. F., 2002). Heart period variability (HPV) has been shown to be an important parameter for mental workload relevant for human-computer interface (HCI) (Iszo, L. et al, 1999). Motivated by the progress in affective computing (Picard, R., 1997), significant research is being done on developing computer interfaces that can detect user affective states through measurement of physiology. In our previous work (Rani, P. et al, 2004) we have shown the relationship between anxiety and several physiological parameters like HRV, facial EMG, skin conductance, blood pulse volume, and peripheral temperature. Prinzel et al. (Prinzel, L. J. et al 2003) have studied the effect of an EEG based adaptive automation on tracking performance and workload. The engagement index calculated in their work is based on the P300 component of the Even-Related Potential (ERP). Kulic et al. discussed their approach to estimate intent for human-robot interaction (Kulic, D., & Croft, E., 2003). They focused on the two aspects of intent namely attention and approval, where attention was measured through gesture recognition and eye gaze tracking and approval was measured through facial expressions and physiological signals. Operator physiological response was also studied by Hanajima et al to investigate the impact of robot motion on operator's HRV and electrodermal activity (Hanajima, N. et al, 2005).

Teleoperation has been the focus of research for many years (Sheridan, T. B., 1992), as researchers have been striving to find ways of dealing with autonomy issues, problems of delays, dexterity and sensory information retrieval. A lot of emphasis has been placed in the design and development of multimodal user interfaces that provide visual and tactile feedback to the operator in order to give a real-life feel. Some researchers have focused on methods aimed at reducing operator workload. For instance Arkin and Ali's work (Ali, K. S., & Arkin, R. C., 2000) deals with behavior-based design of robots that can interact with humans to lessen their workload. Fong et al. in (Fong, T. W. et al, 2001) use collaboration, human-robot dialogue and waypoint-based driving for vehicle teleoperation. These techniques are expected to enable a single operator effectively control multiple mobile robots with his/her limited cognitive resources. However, relatively little work has been done to provide the operator feedback to the robot regarding workload/fatigue/loss of attention. Crandall et al. have experimentally shown that in teleoperation, robot effectiveness is inversely proportional to neglect (Crandall, J. W. & Goodrich, M. A., 2002). That is, the lesser attention an operator gives to the robot during a teleoperation task, the worse will be the robot's performance.

3. Objectives

The main objectives of the paper are to demonstrate: 1) real-time detection of engagement level of an operator can be achieved based on physiological signals; and 2) a teleoperated mobile robot can adapt its behavior according to the level of engagement that it receives from the operator controlling it.

Due to practical difficulties in eliciting loss of engagement with robots in a laboratory environment with limited resources, the experiment to prove the above-mentioned objectives was done in two parts. First, participants who volunteered for the study were engaged in a carefully designed experiment that involved participation in the game of Pong and an Anagram solving task. These tasks were designed to elicit various levels of engagement, boredom, frustration, anger and anxiety from them. The purpose of generating multiple affective states was to allow us to design analysis tools that could detect engagement from a variety of affective states. Second, a teleoperation experiment with the mobile robot, named Oracle (www.arrickrobotics.com), was performed where the physiological data collected in the previous experiment was streamed in continuously to the robot as if it were coming in real-time from the operator. The robot was expected to adapt its behavior to the various levels of operator engagement.

The two-part experiment described above is open loop. It is expected to serve as a proof-of-concept experiment demonstrating the use of physiological feedback to adapt robot behavior. There were several reasons as to why the physiological monitoring and teleoperation were not space and time collocated.

1. Eliciting engagement, frustration and other affective states mentioned earlier through computer tasks is less resource consuming (e.g., each participant was made to play 6 hours of computer games in order to get training and testing data. Doing the same with mobile robots would require longer hours, work area and equipment).
2. Eliciting low engagement in a teleoperation task would require hours of operation as our participants are generally excited about operating a robot and do not show loss of engagement until several sessions of repetitive work.
3. Training participants to operate mobile robots would take additional resources.

However, it is expected that in the future, better availability of resources would enable us to perform field experiments with professional teleoperators. These experiments would be closed-loop, where both the physiological state of the operator working with a robot will be monitored in real-time and the operator responses to change in robot behaviors will be evaluated.

4. Engagement Detection

Two PC based cognitive tasks were designed to elicit several affective states including engagement (or the lack of it) in the participants. Physiological data from participants were collected during the experiment. A part of this data was employed to train the regression-tree based engagement detection system (described in Section 4.3) and the other part to run experiments with the mobile robot, Oracle. A regression tree based prediction system (Rani, P. et al, 2004) was developed to predict the probable level of operator engagement

4.1 Cognitive Tasks

The aim of the tasks was to invoke in the participants the following five affective states: engagement, anxiety, boredom, frustration and anger. The tasks chosen were solving anagrams and playing Pong. The anagram-solving task has been previously employed to explore relationships between both electrodermal and cardiovascular activity with mental anxiety (Pecchinenda, A., & Smith, C. A., 1996). Emotional responses were manipulated in this task by presenting the participant with anagrams of varying difficulty levels, as established through pilot work. A long series of trivially easy anagrams caused boredom, an optimal mix of solvable and difficult anagrams caused engagement, unsolvable or extremely difficult anagrams elicited frustration and giving time deadlines generated anxiety. All these conditions were well tested during the development stage of the task design and piloting.

The Pong task consisted of a series of trials each lasting up to four minutes, in which the participant played a variant of the early, classic video game "Pong". This game has been used in the past by researchers to study anxiety, performance, and gender differences (Brown, R. M. et al, 1997). Various parameters of the game were manipulated to elicit the required affective responses. These included: ball speed and size, paddle speed and size, sluggish or over-responsive keyboard and random keyboard response. Low speeds and large sizes of ball and paddle made games boring after a while, whereas high speed ball and paddle along with smaller sizes of the two made the game engaging. Very high speeds caused anxiety at times. Sluggish or over-responsive keyboard induced frustration and anger. The relative difficulties of various trial configurations were established through pilot work.

During the experiment, participants were presented with cognitive computer tasks that elicited a variety of affective responses. Six participants (four women and two men) took part in the experiment. Their age range was from 24 to 45 years. After initial briefing regarding the computer tasks, sensors were attached to the participant's body. Each participant took part in six sessions of the above two tasks – three one hour sessions of solving anagrams and three one hour sessions of playing Pong. These tasks spanned a period of one month. In each session, before starting the actual tasks baseline recording was done which was used later to offset day-variability. Each session consisted of 3 minute epochs followed by a questionnaire for self-reporting. During the tasks, the participant's physiology was monitored with the help of wearable biofeedback sensors and Biopac data acquisition system (www.biopac.com). The Biopac sensors are small, easy to wear and unobtrusive during the tasks. The digitally sampled sensor information was sent serially to the computer using an Ethernet cable. The signals monitored consisted of electrocardiogram, bio-impedance, electromyogram (from the corrugator, zygomaticus and upper trapezius muscles), galvanic skin response, peripheral temperature, blood volume pulse, and heart sound.

During the tasks, the participants periodically reported their perceived subjective emotional states. This information was collected using a battery of five self-report questions (regarding their perceived anxiety, engagement, anger, frustration and boredom) rated on a nine-point Likert scale. Self-reports were used as reference points to link the objective physiological data to participants' subjective affective state. Each task sequence was subdivided into a series of discrete epochs that were bounded by the self-reported affective state assessments. These assessments occurred every three minutes for the anagram task and every 2-4 minutes for the Pong task. The participants reported their affective state on a scale of 1-9 where 1 indicated the lowest level and 9 indicated the maximum level.

4.2 Physiological Basis and Signal Measurement

There is good evidence that the physiological activity associated with affective state can be differentiated and systematically organized (Bradley, M. M., 2000). The transition from one emotional state to another is generally accompanied by dynamic shifts in indicators of Autonomic Nervous System (ANS) activity. The physiological signals we examined were: various features of cardiovascular activity, including interbeat interval, relative pulse volume, pulse transit time, heart sound, and pre-ejection period; electrodermal activity (tonic and phasic response from skin conductance) and electromyogram (EMG) activity (from corrugator supercilii, zygomaticus, and upper trapezius muscles). These signals were selected because they were likely to demonstrate variability as a function of our targeted affective states as well as they could be measured non-invasively and were relatively resistant to movement artifact. A detailed description of the physiological signals and features can be found in our previous work (Rani, P. et al, 2004). Multiple features as shown in Table 1 were derived for each physiological measure. "Sym" is the power associated with the sympathetic nervous system activity of the heart (in the frequency band 0.04-0.15 Hz.). "Para" is the power associated with the heart's parasympathetic nervous system activity (in the frequency band 0.15-0.4 Hz.). InterBeat Interval (IBI) is the time interval in milliseconds between two "R" waves in the ECG waveform in millisecond. IBI ECGmean and IBI ECGstd are the mean and standard deviation of the IBI. Photoplethysmograph signal (PPG) measures changes in the volume of blood in the fingertip associated with the pulse cycle, and it provides an index of the relative constriction versus dilation of the blood vessels in the periphery. Pulse transit time (PTT) is the time it takes for the pulse pressure wave to travel from the heart to the periphery, and it is estimated by computing the time between systole at the heart (as indicated by the R-wave of the ECG) and the peak of the pulse wave reaching the peripheral site where PPG is being measured. Heart Sound signal measures sounds generated during each heartbeat. These sounds are produced by blood turbulence primarily due to the closing of the valves within the heart. The features extracted from the heart sound signal consisted of the mean and standard deviation of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform. Bioelectrical impedance analysis (BIA) measures the impedance or opposition to the flow of an electric current through the body fluids contained mainly in the lean and fat tissue. A common variable in recent psychophysiology research, pre-ejection period (PEP) derived from impedance cardiogram (ICG) and ECG measures the latency between the onset of electromechanical systole, and the onset of left-ventricular ejection and is most heavily influenced by sympathetic innervation of the heart. Electrodermal activity consists of two main components – Tonic response and Phasic skin conductance response. Tonic skin conductance refers to the ongoing or the baseline level of skin conductance in the absence of any particular discrete environmental events. Phasic skin conductance refers to the event related changes that occur, caused by a momentary increase in skin conductance (resembling a peak). The EMG signal from Corrugator Supercilii muscle (eyebrow) captures a person's frown and detects the tension in that region. It is also a valuable source of blink information and helps us determine the blink rate. The EMG signal from the Zygomaticus Major muscle captures the muscle movements while smiling. Upper Trapezius muscle activity measures the tension in the shoulders, one of the most common sites in the body for developing stress. The useful features derived from EMG activity were: mean, slope, standard deviation, mean frequency and median frequency. Blink movement could be detected from the Corrugator Supercilii activity. Mean amplitude of blink activity and mean interblink interval were also calculated from Corrugator EMG.

Various signal processing techniques such as Fourier transform, wavelet transform, thresholding, and peak detection, were used to derive relevant features from the physiological signals. All these features mentioned above are powerful indicators of the underlying affective state of the person showing this response. We have exploited this dependence of a person's physiological response on affect to detect and identify engagement in real-time

Some of the above-mentioned physiological signals and a few others have been used by researchers in the past to study attention/engagement in people (Fredriskon, M. et al 1986, Pope, A. T. et al, 1995). Spontaneous skin conductance activity has been shown to be associated with task engagement (Pecchinenda, A., & Smith, C. A., 1996). It was also shown in a later work by Kirby and Smith that increased effort and increased memory load caused an increase in skin conductance responses (SCRs) within a trial; on the other hand, increased vigilance was associated with fewer SCRs (Kirby, L. D., & Smith, C. A., 1997). Fairclough and Venables examined a range of psychophysiological responses – electroencephalogram (EEG), electrocardiogram (ECG), Skin Conductance (SC), electro-oculogram (EOG), and respiratory rate and found that psychophysiology predicted a significant proportion of the variance for task engagement (Fairclough, S. H. & Venables L., 2005). Iani et al. found a significant difference in blood pulse volume amplitude between rest and task periods, suggesting that the measure reflected changes in sympathetic activity due to task engagement (Iani, C. et al, 2004).

Their results also indicated that reduced pulse wave amplitude, signaling vasoconstriction, resulted when participants spent effort. Many of these studies provide a basis for our investigation. However, unlike our work, none of the above-mentioned studies attempts to infer the underlying psychological state in real-time. Furthermore, none of these works is in the field of human-robot interaction.

Various methods of extracting physiological features exist but efforts to identify the exact markers related to emotions, such as anger, fear, or sadness have not been successful chiefly due to person-stereotypy and situation-stereotypy

(Lacey, J. L. & Lacey, B. C., 1958). That is, within a given context, different individuals express the same emotion with different characteristic response patterns (person-stereotypy). In a similar manner, across contexts the same individual may express the same emotion differentially, with different contexts causing characteristic responses (situation-stereotypy). The novelty of the presented affect-recognition system is that it is both individual- and context-specific in order to accommodate the differences encountered in emotion expression. It is expected that in the future with enough data and understanding, affect recognizers for a class of people can be developed.

The participants periodically reported their perceived subjective emotional states. This self-report was collected using a battery of fourteen questions rated on nine-point Likert scales. These questions asked them to report their emotions such as anxiety, challenge, and engagement as related to the task that they just performed. An engagement index was determined from the rating that the participants provided regarding their engagement during the task.

4.3. Engagement Prediction based on Regression Tree

In the previous research works in emotion recognition, change in emotion has been considered either along a continuous dimension (e.g., valence or arousal) or among discrete states. Various machine learning and pattern recognition methods have been applied for determining the underlying affective state from cues such as facial expressions, vocal intonations, and physiology. Fuzzy logic has been employed for emotion recognition from facial expression (Moriyama, T. et al 1999). Fuzzy logic has also been used to detect anxiety from physiological signals by our research group 0 and by Hudlicka et al. in (Hudlicka, E., & McNeese, M. D., 2002). There are several works on emotion detection from speech based on k-nearest neighbors algorithm (Petrushin, V. A., 2000), linear and nonlinear regression analysis (Rani, P. et al 2003). Discriminant analysis has also been used to detect discrete emotional states from physiological measures (Ark, W. et al 1999). A combination of Sequential Floating Forward Search and Fisher Projection methods was presented in (Vyzas, E., & Picard, R. W., 1998) to analyze affective psychological states. Neural networks have been extensively used in detecting facial expression (Zhao, J., & Kearney, G., 1996), facial expression and voice quality (Fellenz, W., A. et al, 2000). The Bayesian approach to emotion detection is another important analysis tool that has been used successfully. In (Qi, Y., & Picard, R., 2002) a Bayesian classification method was employed to predict the frustration level of computer users based on pressure signals from mouse sensors. A Naïve Bayes classifier was used to predict emotions based on facial expressions (Sebe, N. et al, 2002. A Hidden Markov Model based emotion detection technique was investigated for emotion recognition (Cohen, I. et al, 2000).

In this paper we have used regression trees (Breiman, L., 1993) (also known as decision trees) to determine a person's affective state from a set of features derived from physiological signals. The choice of regression tree method emerges from our previous comparative study of four machine learning methods- K-Nearest Neighbor, Regression Tree, Bayesian Network and Support Vector Machine as applied to the domain of affect recognition (Liu, C. et al 2005). The results showed that regression tree technique gave the second best classification accuracy – 83.5% (after Support Vector Machines that showed 85.8% accuracy) and was most space and time efficient. Regression tree method has not been employed before for physiology-based affect detection and recognition. Person-specific regression trees were created to handle the occurrence of person-stereotypy.

Physiological Response	Features derived	Label usedir	Unit of measurement
Cardiac activity	Sympathetic power	Sym	Unit/Square Second
	Parasympathetic power	Para	Unit/Square Second
	Ratio of Sympathetic to Parasympathetic power	Sym Para	Unit/Square Second
	Mean IBI	IBI ECG_{mean}	Milliseconds
	Std. of IBI	IBI ECG_{std}	Standard Deviation (no unit)
	Mean amplitude of the peak values of the PPG signal (Photoplethysmogram)	PPG $Peak_{mean}$	Micro Volts
	Standard deviation (Std.) of the peak values of the PPG signal	PPG $Peak_{std}$	Standard Deviation (no unit)
	Mean Pulse Transit Time	PTT_{mean}	Milliseconds
Heart Sound	Mean of the 3rd,4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Mean d3 Mean d4 Mean d5	No unit
	Standard deviation of the 3rd,4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Std d3 Std d4 Std d5	No unit
Bioimpedance	Mean Pre-Ejection Period	PEP_{mean}	Milliseconds
	Mean IBI	IBI ICG_{mean}	Milliseconds
Electrodermal activity	Mean tonic activity level	$Tonic_{mean}$	Micro-Siemens
	Slope of tonic activity	$Tonic_{slope}$	Micro-Siemens/Second
	Mean amplitude of skin conductance response (phasic activity)	$Phasic_{mean}$	Micro-Siemens
	Maximum amplitude of skin conductance response (phasic activity)	$Phasic_{max}$	Micro-Siemens
	Rate of phasic activity	$Phasic_{rate}$	Response peaks/Second
Electromyographic activity	Mean of Corrugator Supercilii activity	Cor_{mean}	Micro Volts
	Std. of Corrugator Supercilii activity	Cor_{std}	Standard Deviation (no unit)
	Slope. of Corrugator Supercilii activity	Cor_{slope}	Micro Volts/Second
	Mean Interbeat Interval of blink activity	IBI $Blink_{mean}$	Milliseconds
	Mean amplitude of blink activity	Amp $Blink_{mean}$	Micro Volts
	Standard deviation of blink activity	$Blink_{std}$	Standard Deviation (no unit)
	Mean of Zygomaticus Major activity	Zyg_{mean}	Micro Volts
	Std. of Zygomaticus Major activity	Zyg_{std}	Standard Deviation (no unit)
	Slope. of Zygomaticus Major activity	Zyg_{slope}	Micro Volts/Second
	Mean of Upper Trapezius activity	$Trap_{mean}$	Micro Volts
	Std. of Upper Trapezius activity	$Trap_{std}$	Standard Deviation (no unit)
	Slope. of Upper Trapezius activity	$Trap_{slope}$	Micro Volts/Second
	Mean and Median frequency of Corrugator, Zygomaticus and Trapezius	$Zfreq_{mean}$ $Cfreq_{median}$ $Tfreq_{mean}$ etc.	Hertz
Temperature	Mean temperature	$Temp_{mean}$	Degree Centigrade
	Slope of temperature	$Temp_{slope}$	Degree Centigrade/Second

Table 1. Physiological Indices and the Features

Regression tree learning, a frequently used inductive inference method, approximates discrete valued functions that adapt well to noisy data and are capable of learning disjunctive expressions. For the regression tree-based affect recognizer that we built, the input consisted of the physiological feature set and the target function consisted of the affect levels (participant's self-reports that represented the participant's assessment of his/her own affective state). The main challenge was the complex nature of the input physiological data sets. This complexity was primarily due to the (i) high dimensionality of the input feature space (there are currently forty-six features and this will increase as the number of affect detection modalities increases.), (ii) mixture of data types, and (iii) nonstandard data structures. Additionally, a few physiological data sets were noisy where the biofeedback sensors had picked up some movement artifacts. These data sets had to be discarded, resulting in the missing attributes.

In this work, we used regression tree to determine a participant's level of engagement fro his/her physiological response. The objective was to train a regression tree such that it could accept as input an array of feature vectors derived from physiological signals and provide as input the engagement level associated with it. The output was a discrete number in the range 0-9 where 0 signified lowest level and 9 signified highest level of engagement. The steps involved in building a regression tree are shown in Fig. 1. Physiological signals recorded from the participant engaged in PC-based task were processed to extract the input feature set. The participant's self-report at the end of each epoch regarding his/her affective states provided the target variable or the output. While creating the tree, two primary issues were: (i) Choosing the best attribute to split the examples at each stage, and (ii) Avoiding data over fitting. Many different criteria could be defined for selecting the best split at each node. In this work, Gini Index (Breiman, L., 1993) function was used to evaluate the goodness of all the possible split points along all the attributes. For a dataset D consisting of n records, each belonging to one of the m classes, the Gini Index can be defined as:

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^{\,2} \qquad (1)$$

where pi is the relative frequency of class i in D. If D is partitioned into two subsets D1 and D2 based on a particular useful attribute, the index of the partitioned data Gini(D,C) can be obtained by:

$$Gini(D,C) = \frac{n_1}{n} Gini(D_1) + \frac{n_2}{n} Gini(D_2) \qquad (2)$$

where n1 and n2 are the number of examples of D1 and D2, respectively, and C is the splitting criterion. Here the attribute with the minimum Gini Index was chosen as the best attribute to split. We used the Statistics Toolbox of Matlab for regression tree functions such as generating, pruning and evaluating regression trees (www.mathworks.com). Trees were pruned based on an optimal pruning scheme that first pruned branches that gave the least improvement in error cost. Pruning was performed to remove the redundant nodes as bigger, overfitted trees have higher misclassification rates. Thus, based on the input set of physiological features described earlier, the affect recognizer provided a quantitative understanding of the person's affective states.

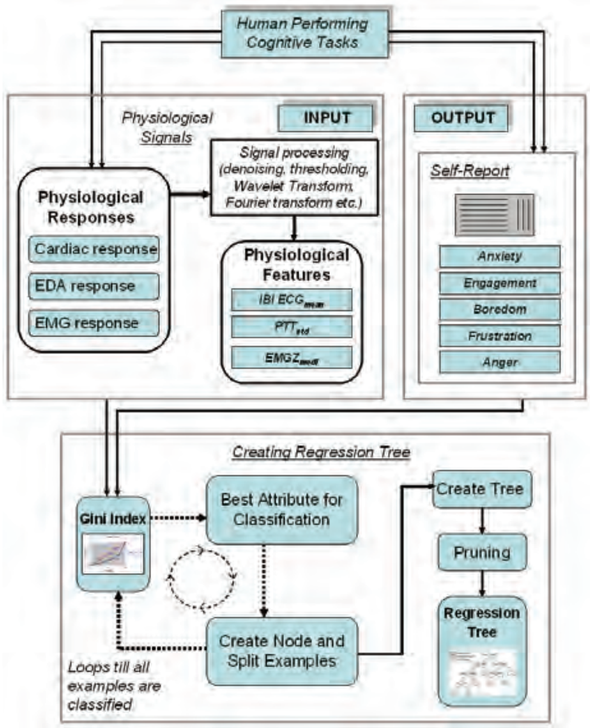

Fig. 1. Creating a Regression Tree

5. Engagement-Based Teleoperation

5.1 Control Architecture for Implicit Human-Robot Interaction

For a human-robot interaction to mimic similar human-human interaction, it is essential that both the robot and the human have implicit as well as explicit communication with each other. This requires a systematic information flow between the human, robot and the environment. A generalized model of human-machine system developed by Riley (Riley, V., 1989) represents such an information flow that can be systematically modified according to any rule-base to represent a particular level of automation in human-machine interaction. It is a powerful front-end analysis method that can be employed to identify human-machine protocols as well as the automation concerns that accompany the design of such systems. The general model represents the most complex level of automation embedded in the most complicated form of human-machine interaction. We modified this model to represent the specific system developed for human-Oracle interaction in a teleoperation experiment. This system mimicked a teleoperation task where an operator teleoperated a mobile robot in a given workspace. In this case Oracle is expected to behave as an intelligent partner to the operator. This requires Oracle to respond appropriately to the engagement levels of the operator while not undermining the importance of its own safety and goals. The reduced architecture is shown in Figure 1.

As seen in Figure 2, in the top-left "robot input" quadrant, Oracle receives information from both the world and the operator through various sensors. The world information is obtained through the infrared sensors, touch sensors, light sensors etc. Oracle receives this information, infers the world state and forms a model of the world for itself. Operator information that is received by Oracle is of two types – control input and affective input. The control input contains teleoperation information and is provided to Oracle through a joystick. The affective information is obtained through biofeedback sensors that measure physiological signals of the operator, for instance- cardiac activity, electrodermal activity, muscle activity, temperature etc. This information is used to interpret the affective state of the operator, for example to judge whether the operator is getting inattentive, fatigued, or stressed (implicit communication). Oracle also receives explicit commands from the operator through the control sensors (explicit communication). Instead of blindly following every teleoperation command of the operator, Oracle continuously evaluates the engagement level of the operator and modifies its behavior accordingly. On detecting low levels of engagement, Oracle alerts the operator and in some cases can even stop itself.

The “robot output” quadrant contains nodes that determine the behavior of Oracle. Oracle uses its representation of the world, knowledge of its own goals and the urgency level of the situation to determine the best course of action. Oracle is assisted by a set of rules to determine its behavior as shown in Figure 3. In our previous works 0,0 we have employed a hybrid subsumption architecture model to determine Oracle’s behavior. Oracle’s decision results either in an action or initiation of speech on its part. In the “human input” quadrant the human receives information both from the world and Oracle. The human can perceive the dialogues initiated by Oracle and observe its behavior. He/she then exploits this knowledge to infer Oracle’s state as well as predict what it might do next. This inference, the world representation that the human forms, and his/her own ultimate goals are all employed by the human to determine the next action. The resulting human’s action might be simply to monitor Oracle’s actions or issue a command to it. Hence, in each cycle of the loop, there is a methodical information flow in between the world, Oracle and the human. At the very fundamental level, this is a sense-infer-plan-act loop wherein, Oracle and the human utilize the available information to interact with each other and to take actions that influence each other and the world.

5.2 Teleoperation Experiment

The objective of the experiment was to develop and implement real-time, affect-sensitive human-robot co-ordination during teleoperation. The robot was expected to recognize and respond to the varying levels of engagement of the human operator.

However, due to the limitations explained in Section III, the experiment was performed in two parts. First, we designed a regression tree based affect-recognizer that could identify or “understand” the physiological responses of a person. Second, we implemented a human-robot co-operation task consisting of Oracle being teleoperated by an operator/user whose (apparent) physiological state was being continuously monitored. The first part of the experiment is described in Section IV. The second part of the experiment consisted of implementing a real-time human-robot interaction during teleoperation. Oracle was supposed to recognize the operator’s level of engagement through continuous physiological sensing, and act accordingly to address the situation. Since closed-loop experiment could not be performed, the physiological signals collected in the first part were streamed in at real-time speed to Oracle as if it were coming from the person controlling the robot. This allowed us to demonstrate that if such physiological signals are generated during a human-robot task, the robot could detect the engagement level of the operator and subsequently modify its behaviors. Controlling the Oracle was done by sending low-level commands from a desktop PC using a standard RS-232 serial port. Commands were wirelessly sent to Oracle by the operator that allowed control over the drive motors, heading motion, gripper, sensors, etc. The teleoperation task consisted of picking and depositing of objects (coke cans) by Oracle in a given workspace. Oracle was teleoperated by an operator via a joystick in accomplishing this task. While the teleoperation task continued, the physiological data collected in the first part of the experiment was sent to the robot as if coming from the operator in real time. The behavior modifications of Oracle depended upon the level of operator engagement detected. The continuous values obtained from regression tree were discretized to obtain three levels of engagement – high, medium and low. As shown in Figure 3 Oracle could switch between three types of behavior depending upon the level of operator engagement. When it detected high levels of engagement, it followed the operator’s commands being sent to it. As the level of engagement decreased to medium level, it alerted the operator by speaking out its current movement, for instance, “I am turning left now.” When the level of engagement dipped lower, Oracle lowers its speed to a minimum after giving a warning to the operator and eventually stopped.

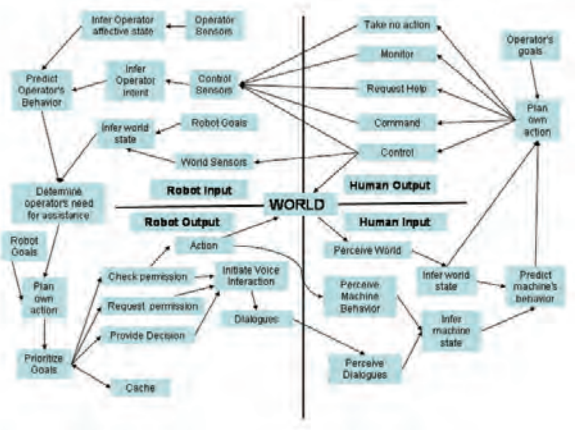

Fig. 2. Human-Robot Interaction Framework (Modified From Riley’s Original Information Flow Model)

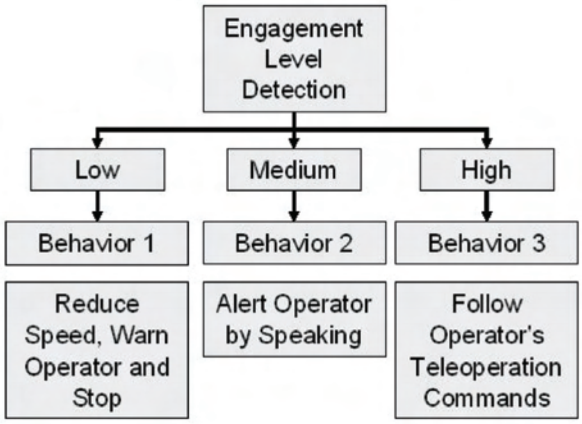

Fig. 3. Behavior Types for Mobile Robot

6. Results

6.1 Engagement Detection

Forty nine features/indices were derived from the following measured physiological signals – ECG, EMG (corrugator, zygomaticus, and upper trapezius), peripheral temperature, galvanic skin response, blood pulse volume, heart sounds, and. bio-impedance. On determining the correlation between the indices derived from the physiological signals and the self reported level of engagement, it was found that certain physiological indices were well correlated with the state of engagement. A correlation of 0.3 and above was considered a significant value and was supported by previous research. Fig. 4 shows the correlation values for indices that were highly correlated with engagement for Participant 5. The useful indices for this participant were IBI - dz/dt (Interbeat Interval from the difference signal of bio-impedance), IBI- ECG (R-R interval from ECG), Tonic Mean (mean level of tonic skin conductance), Phasic Rate, Phasic Mean, Phasic Max (rate, mean amplitude and maximum amplitude of the phasic response of skin conductance), Temp std (standard deviation of the peripheral skin temperature), blink amp (mean amplitude of the eye blink response), and Zyg Med Freq (Median frequency of zygomaticus EMG).

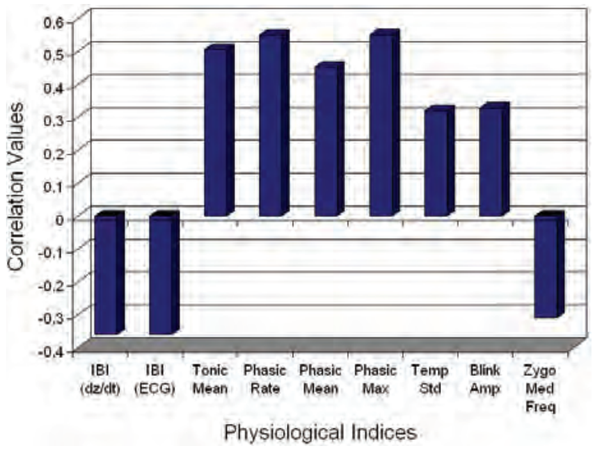

Fig. 4. Correlation of Physiological Indices with Self-Reported Engagement for Participant #1

For each of the six participants there were distinct physiological indices that showed significant correlation with their state of engagement. However, these indices were not identical for all participants due to the well-known phenomenon of subject-stereotypy (Within a given context, different individuals express the same emotion with different characteristic response patterns). Fig. 5 shows the correlations for Participant #1. Person-stereotypy can be seen on comparing the two plots.

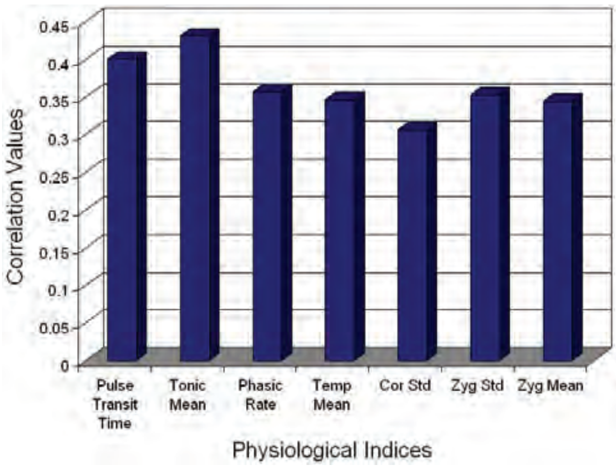

Fig. 5. Correlation of Physiological Indices with Self-Reported Engagement for Participant #5

Not only were the useful indices different for the two participants, the extent of correlation for the same index is different for each participant. It was observed that phasic rate was well correlated for all the participants. The other important indices were mean pulse transit time, mean blink amplitude, mean tonic level, and standard deviation of zygomaticus EMG. Such individual differences can be accounted for by the phenomena of person-stereotypy. We used regression tree methodology to compute a participant's level of engagement from a set of features derived from his/her physiological response. Our approach was to determine the specific pattern for each person and automatically create a regression tree according to individual responses.

It was also observed that based on the indices, it was possible to distinguish between different affective states, for instance between engagement and anxiety and between engagement and boredom. Due to space limitation, we only present the results of selected participants, even though the traits observed were generic. Fig. 5 shows that the indices that indicate engagement can be differentiated well from the state of boredom while Fig. 6 shows that engagement is also well differentiated from the state of anxiety for Participant #5. All the six participants showed clear distinction between the five measured affective states based on physiological indices.

6.1 Experimental Demonstration of Adaptive Robot Behavior

During the teleoperation experiment, an operator would control Oracle via a joystick to perform a pick and place task. Oracle was supposed to pick specific objects (empty coke cans) from the workspace and place them in a specified area. At the same time, Oracle also received physiological feedback as if it were coming in real-time from the operator. Oracle modified its behaviors depending upon the perceived engagement level of the operator (Fig. 3). During the experiment, an engagement index inferred from the various physiological indices described before was computed every 30 seconds. The engagement index was computed using the regression tree that was previously created from the data collected during the PC based cognitive tasks. Self-reports of participants were available every three minutes to validate the correctness of the real-time algorithm used for engagement detection. There were 36 teleoperation trials done (six for each participant). Each session lasted one hour. Fig. 8 shows a teleoperation trial in progress.

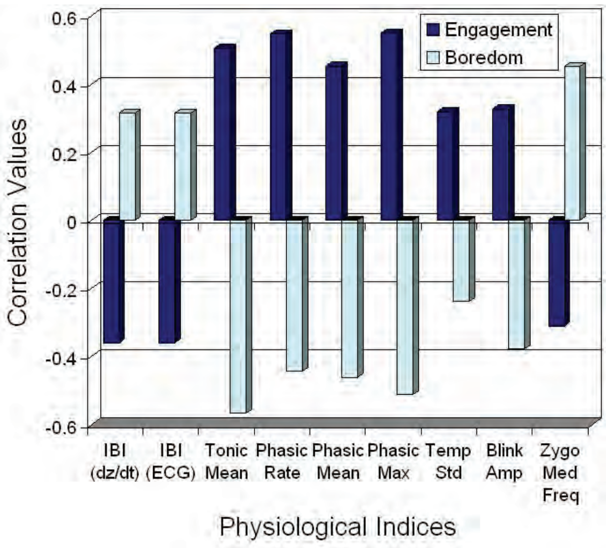

Fig. 6. Comparison between Correlation of Self-Reported Boredom and Self-Reported Engagement with Physiological Indices for P #5

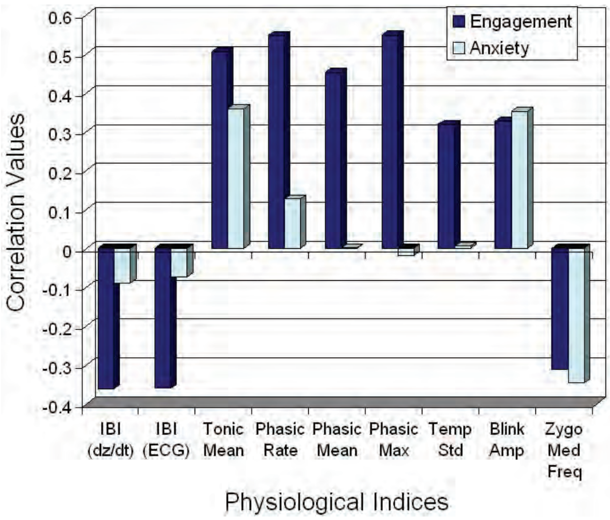

Fig. 7. Comparison between Correlation of Self-Reported Anxiety and Self-Reported Engagement with Physiological Indices for P #5

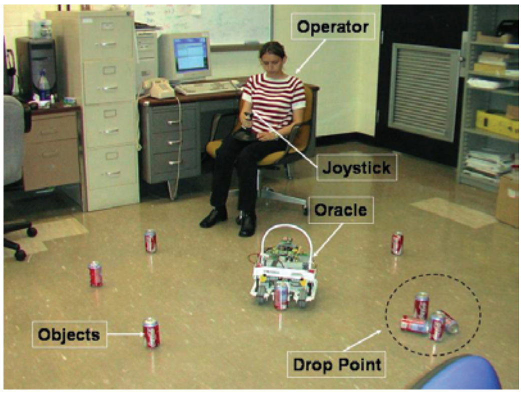

Fig. 8. Teleoperation Experiment

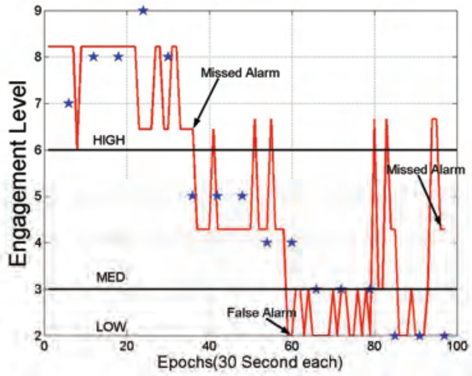

Fig. 9. Continuous Evaluation of Engagement Index

Fig. 9 shows the continuous evaluation of the engagement level in real-time along with the periodic self-reporting from Participant #1.

The solid line indicates the predicted value of engagement (calculated every 30 seconds) and the stars indicate the periodic self-reporting (every 3 minutes) of the participant regarding his/her perceived engagement. It can be seen that during the one-hour trial, there were two missed alarms and one false alarm.

For Participant #3, a confusion matrix (Table 2) shows the number of false alarms and missed alarms during the real time detection of engagement from the physiological data during all of the six sessions (training data set size is 50). The engagement values have been divided into three different groups – low, medium and high according to the rule based system that determines the behavior changes. It can be seen that the detection is accurate (17+47+20)/97=86.60 % times, alarms are missed 7/97=7.22 % and false alarms occur 6/97=6.19 % times.

Actual Value \ Predicted Value	Low	Medium	High
Low	17	4	1
Medium	1	47	2
High	0	5	20

Table 2 Confusion Matrix for P #3

Fig. 10 shows the variation in the percentage accuracy of the engagement detection system with change in the size of the training set across all participants. The training datasets were formed by selecting a subset of the entire training dataset while keeping the testing data set constant. The accuracies presented in Fig. 10 were computed over the invariant testing dataset.

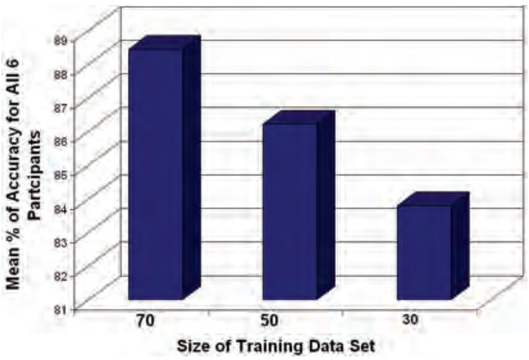

Fig. 10. Mean % Accuracy with Change in Size of Training Data Set For All Six Participants

It can be seen that as the size of training data set was increased the overall accuracy became higher. We expect that in the future we can train our engagement detection system on larger training datasets and thereby achieve higher accuracy in detecting engagement.

7. Conclusions and Future Work

A proof-of concept teleoperation experiment demonstrating detection of operator engagement level by the robot was presented. An innovative human-robot interaction structure was designed and developed wherein the robot was sensitive to the engagement of the operator it worked with and could adapt its behavior according to this perception. This approach synergistically combined concepts in affective computing, psychology, and robotics to develop a robotic system capable of combining implicit and explicit channels of communication from the human to intelligently determine its optimal behavior. A regression tree based prediction system yielded reliable engagement prediction with approximately 88% success. We experimentally demonstrated that the robot could adapt its behavior based on the engagement level of the operator as determined from his/her recorded physiological signals that were sent to the robot in real-time.

Future work would consist of expanding the range of tasks and contexts to which this framework can be applied and increasing the reliability and robustness of engagement detection. The next step would be to test the presented human-robot interaction framework in closed-loop experiment. We would also like to work towards increasing the range of affective states detected beyond engagement to include frustration, fatigue, boredom, anxiety and anger.

8. References

Ali, K. S. & Arkin, R. C. (2000). Mutiagent Teleautonomous Behavioral Control, In: Machine Intelligence and Robotics Control, Vol.1, No. 2, pp. 3-10, 2000

Ark, W., Dryer, D. & Lu, D. (1999). The Emotion Mouse, Human-Computer Interaction: Ergonomics and User Interfaces 1, Bullinger. H. J. and J. Ziegler (Eds.), Lawrence Erlbaum Assoc., London. pp. 818-823, 1999

Bradley, M. M. (2000). Emotion and motivation," In: Handbook of Psychophysiology, J.T. Cacioppo, L. G. Tassinary, and G. Berntson (Eds.) pp. 602-642, University Press, New York

Breiman, L. (1993). Classification and Regression Trees, Chapman & Hall, Boca Raton

Brown, R. M., Hall, L. R., Holtzer, R., Brown, S. L., & Brown, N.L. (1997). Gender and Video Game Performance, Sex Roles, Vol. 36, No. 11-12, pp. 793– 812, June 1997

Cohen, I., Garg, A. & Huang, T. S. (2000). Emotion recognition using multilevel HMM, Proceedings of NIPS Workshop on Affective Computing, Colorado, 2000

Crandall, J. W. & Goodrich, M. A. (2002). Characterizing efficiency of human robot interaction: a case study of shared-control teleoperation, IEEE/RSJ International Conference on Intelligent Robots and System, pp. 1290-1295, Switzerland, 2002

Drury, J. L., Scholtz, J. & Yanco, H. A. (2003). Awareness in Human-Robot Interactions, Proceedings of the IEEE Conference on Systems, Man and Cybernetics, Washington DC., October 2003

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement, Proceedings of the Human Factors Society 32nd Annual Meeting, Vol. 1, pp. 97–101, California, 1988

Fairclough, S. H. & Venables L., 2005. Psychophysiological predictors of task engagement and distress, Human Factors in Design, Safety, and Management In D. de Waard, K.A. Brookhuis, R. van Egmond, and Th. Boersema (Eds.), pp. 349 – 362, Maastricht, the Netherlands: Shaker Publishing

Fellenz, W. A., Taylor, J. G., Cowie, R., Douglas-Cowie, E., Piat, F., Kollias, S., Orovas, C. & Apolloni, B. (2000) On emotion recognition of faces and of speech using neural networks, fuzzy logic and the ASSESS systems, Proceedings of IEEE-INNS-ENNS International Joint Conference on Neural Networks, pp: 2093, 2000

Fong, T. Grange, S., Thorpe, C. & Baur, C. (2001). Multi-robot remote driving with collaborative control, 10th IEEE International Workshop on Robot and Human Interactive Communication, 2001

Fong, T. W. (2001). Collaborative Control: A Robot-Centric Model for Vehicle Teleoperation, Doctoral dissertation, Tech. report CMU-RI-TR-01-34, Robotics Institute, Carnegie Mellon University, November 2001

Fredrikson, M., Gunnarsson, R. & K. Klein. (1986). Emotion, attention, and autonomic arousal, Biological Psychology, Vol. 23, No. 1) pp. 96-97, August 1986

Hanajima,N., Ohta, Y., Hikita, H. & Yamashita, M. (2005). Investigation of impressions on approach motion of a mobile robot based on psychophysiological analysis, ROMAN, Nashville, USA, August 2005

Hayakawa, Y. & Sugano, S. (1998). Real time simple measurement of mental strain in machine operation, ISCIE 1998 Japan-U.S.A. symposium on Flexible Automation, pp. 35-42, Japan, 1998

http://www.biology-online.org/dictionary/engagement
http://www.biopac.com/
Hudlicka, E. & McNeese, M. D. (2002). Assessment of user affective and belief states for inference adaptation: application to an air force pilot task, User Modeling and User Adapted Interaction, Vol. 12, pp. 1-47
Iani, C., Gopher, D. & Lavie, P. (2004). Effects of task difficulty and invested mental effort on peripheral vasoconstriction, Psychophysiology, Vol. 41, No. 5, pp. 789, September 2004
Iszo, L., Mischinger, G. & Lang, E. (1999). Validating a new method for ergonomic evaluation of human computer interfaces, Periodica Polytechnica Ser. Soc. Man. Sci., Vol. 7, No. 2, pp. 119-134, 1999
Kirby,L. D. & Smith, C. A. (1997) Spontaneous skin conductance activity as a marker of task engagement: Refining the relations, Psychophysiology, Vol. 34, S53, 1997
Kramer, A. F., Sirevaag, E. J. & Braune, R. (1987). A Psychophysiological assessment of operator workload during simulated flight missions, Human Factors. Vol. 29, No. 2, pp. 145-160, 1987
Kulic, D. & Croft, E. A. (2003). Intent Based Control for Human Robot Interaction, International Conference on Advanced Robotics, Portugal, June 2003
Lacey, J. L. & Lacey, B. C. (1958). Verification and extension of the principle of autonomic response-stereotypy, American Journal of Psychology, Vol. 71, No. 1, pp. 50-73, 1958
Liu, C., Rani, P. & N. Sarkar (2005). Comparison of Machine Learning Techniques for Affect Detection in Human Robot Interaction, IEEE/RSJ International Conference on Intelligent Robots and Systems, August 2005, Canada
Moriyama, T., Saito, H. & Ozawa, S. (1999). Evaluation of the relation between emotional concepts and emotional parameters on speech, IEICE Journal, J82-DII(10), pp. 1710-1720, 1999
O'Hare, D., Wiggins, M., Batt, R., & Morrison, D. (1994). Cognitive failure analysis for aircraft accident investigation, Ergonomics, Vol. 37, pp. 1855-69
Pecchinenda, A. & Smith, C. A. (1996). The motivational significance of skin conductance activity during a difficult problem-solving task, Cognition and Emotion, Vol. 10, pp. 481-503, 1996
Petrushin, V. A. (2000) Emotion recognition agents in real world, AAAI Fall Symposium on Socially Intelligent Agents: Human in the Loop
Picard, R. (1997). Affective Computing, The MIT Press, Cambridge, Massachusetts
Pope, A. T., Bogart, E. H & Bartolome, S. (1995). Biocybernetic system evaluates indices of operator engagement, Biological Psychology, Vol. 40, pp. 187-196, 1995
Prinzel III, L. J., Freeman, F. G., Scerbo, M. W., Mikulka P. J & Pope A. T. (2003). Effects of a Psychophysiological System for Adaptive Automation on Performance, Workload, and the Event-Related Potential P300 Component, Human Factors, Vol. 45, No. 4, Winter 2003, pp: 601-613
Qi, Y. & Picard, R. W. (2002). Context-sensitive Bayesian Classifiers and Application to Mouse Pressure Pattern Classification, Proceedings of International Conference on Pattern Recognition, Quebec City, Canada, 2002
Rani, P. Sarkar, N., Smith, C. A. & Kirby, L. (2004). Anxiety Detecting Robotic Systems – Towards Implicit Human-Robot Collaboration, Robotica, Vol. 22, No. 1, 2004, pp. 85-95.
Rani, P., Sarkar, N. & Smith, C. A. (2003). Affect-Sensitive Human-Robot Cooperation –Theory and Experiments, IEEE International Conference on Robotics and Automation, pp: 2382-2387, Taiwan, September 2003
Rani, P., Sarkar, N., & Smith, C. A. (2003) Affect-Sensitive Human-Robot Cooperation –Theory and Experiments, IEEE International Conference on Robotics and Automation, pp. 2382-2387, Taiwan, September 2003
(Accepted) Rani, P., Sarkar, N., Smith, C. A., & Adams, J. A. (2005). Affective Communication for Implicit Human-Machine Interaction, Journal of Advanced Engineering Informatics
Reason, J. (1990). Human Error, Cambridge University Press
Riley, V. (1989). A General Model of Mixed-Initiative Human-Machine Systems, Proceedings of the Human Factors Society 33rd Annual Meeting, pp. 124-129
Schafer, R. (2003). Robotics to play major role in future warfighting, (United States Joint Forces Command), Available from: http://www.jfcom.mil/newslink/storyarchive/2003/pa072903.htm, Accessed: 2006-09-04
Sebe, N., Cohen, I., Garg, A. & Huang, T. S. (2002). Emotion Recognition using a Cauchy Naive Bayes Classifier, Proceedings of International Conference on Pattern Recognition, pp: 10017-10021, 2002
Shappell, S. & Wiegmann, D. (1996). U. S. Naval Aviation mishaps 1977-92: Differences between single- and dual-piloted aircraft, Aviation, Space, and Environmental Medicine, Vol. 67, pp. 65-9, 1996
Sheridan, T. B. (1992) Telerobotics, Automation, and Human Supervisory Control, The MIT Press
Treat, J. R., McDonald, N. S., Shinar, D., Hume, R. D., Mayer, R. E., Stansifer, R., L., & Castellan, N. J. (1977). Tri-Level Study of the Causes of Traffic Accidents, Causal Factor Tabulations and Assessment, Vol. 1, Report No. DOT-HS-805-085, National Highway Traffic Safety Administration
Vyzas, E., & Picard, R. W. (1998). Affective Pattern
Classification, AAAI Fall Symposium Series: Emotional and Intelligent: The Tangled Knot of Cognition, Orlando, Florida, October 1998
Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple Psychophysiological measures, The International Journal of Aviation Psychology, Vol. 12, No. 1, p. 3-18, 2002
www.arrickrobotics.com
www.mathworks.com
Zhao, J. & Kearney, G, Classifying facial movement by backpropagation neural networks with fuzzy inputs, Proceedings of International Conference on Neural Information Processing, pp. 454-457, 1996