Stress Recognition Using Non-invasive Technology

Jing Zhai [1] and Armando Barreto[1,2]

[1]Electrical and Computer Engineering Department
[2]Biomedical Engineering Department
Florida International University
{ jzhai002, barretoa }@fiu.edu

Abstract

The need to provide computers with the ability to estimate the affective state of their users is a major requirement for the practical implementation of Affective Computing concepts. This research aims at sensing and recognizing typical negative emotional states, especially "stress", when the user is interacting with the computer. An integrated hardware – software setup has been developed to achieve automatic assessment of the affective status of a computer user. A computer-based "Paced Stroop Test" is designed to act as a stimulus to elicit emotional stress in the subject. Four signals: Blood Volume Pulse (BVP), Galvanic Skin Response (GSR), Pupil Diameter (PD) and Skin Temperature (ST) are monitored and analyzed to differentiate affective states in the user. Several signal processing techniques are applied to the signals collected to extract the most relevant features in the physiological responses and feed them into learning systems, to accomplish the affective state classification. Three learning algorithms are applied to this classification process and their performance is compared. Results indicate that the physiological signals monitored do, in fact, have a strong correlation with the changes in emotional state of our experimental subjects.

Introduction

Hudlicka (Hudlicka, 2003) and others have described the importance of the emotional and affective factors in human-computer interaction. Reeves and Nass (Reeves & Nass, 1996) have proposed that the human-computer interaction should follow the patterns of human-human interactions. Furthermore, the emerging discipline of Affective Computing (Picard, 1997) has illuminated the potential that could be reached by computer systems endowed with affective understanding. Fulfillment of that promise, however, requires a robust mechanism to make the computer aware of the affective state of the user. Our research attempts to detect the emotional state identified with 'stress' of the computer users, through four physiological signals that can be measured non-invasively and non-intrusively. This paper outlines our pursuit of that goal, by application of digital signal processing and machine learning tools.

Physiological Responses to Affective Stimuli

The autonomic nervous system (ANS) of human beings comprises three separate sub-systems, the sympathetic, parasympathetic, and enteric divisions. Knowledge about the basic structure and function of the ANS is an essential pre-requisite for the analysis of psychophysiological phenomena. The parasympathetic division predominates under resting conditions and the sympathetic division activates during periods of exertion, stress, or emergency. ANS effector organs, such as the heart and the sweat glands, are strongly influenced by these divisions and this influence is the conduit by which ANS activity mediates stress responses and emotional arousal, which are two core concepts of the mind-body interface as well as many psychosomatic diseases (Hugdahl, 2001).

Several previous reports in the literature on Affective Computing involve monitoring of the Blood Volume Pulse (BVP), the Galvanic Skin Response (GSR) and the Heart Rate Variability (HRV) (Picard, 2001; Rani et al., 2003; Scheirer et al., 2002). Variations of the Pupil Diameter (PD) have also been investigated during and after auditory emotional stimulation (Partala & Surakka, 2003) and the finger tip Skin Temperature (ST) has been reported as an indicator for sympathetic responses (Kistler, Mariauzouls & von Berlepsch, 1998). However, the concurrent analysis of the variations of pupil diameter and skin temperature with other physiological signals to indicate the emotional state of a subject has not been fully investigated.

For our research, four sensors were chosen to measure physiological signals that could give a continuous reading and could be collected in a minimally invasive fashion from the computer users. These sensors measured skin resistance, heart activity, the pupil diameter and the skin temperature. The ultimate goal of the system designed in our research is to provide continuous monitoring and digital signal processing of physiological variables to inform a computer system of its user's affective state.

Experimental Setup Design

Thirty two healthy subjects recruited from the student body at Florida International University (ages 21 – 42) participated in this study. To assure reliability of the changes of PD and ST, the lighting and the temperature of the environment were kept at a constant level during the experiments and the illumination of the Eye-gaze system was the same for all the subjects. A GSR2 module, by Thought Technology LTD (West Chazy, New York) was used to collect the subject's Galvanic Skin Response (GSR) and photoplethysmography (PPG) was used to measure the blood volume in the skin capillary bed of the left ring finger. The subject's skin temperature was measured with an LM34 IC that provides a linear output between -50 and 300 degrees Fahrenheit. Ordinarily, the finger tip temperature is expected to be between 75^{o}F and 100 ^{o}F, unless vasoconstriction due to sympathetic response (caused by pain or mental stress) occurs (Kistler et al., 1998). So the output of the temperature sensor was buffered and scaled by a differential amplifier (with a gain of 31V/V) to achieve an analog signal appropriate for digitization. The temperature sensor was attached to the distal phalanx of the left thumb finger with the help of Velcro. The sampling rate for GSR, BVP and ST recording was set to 360 samples/sec. To get an accurate and continuous pupil diameter signal, we used the ASL-504 eye gaze tracking system running at a different sampling rate of 60 samples/sec and later upsampled the signal to 360 samples/sec to match the common sampling rate of the GSR, BVP and ST recording system. The detailed description of the GSR and BVP sensors, and the signal collection and associated synchronization method can be found in our previous report (Barreto & Zhai, 2003).

In order to measure the changes that take place in these four signals when stress sets in, a hardware / software system has been developed to:

a) Provide an appropriate stimulus, capable of eliciting stress in the subjects participating in the experiment;
b) Provide synchronization signals for the rest of the instrumental setup, so that the segments of the several signals recorded under the stress state can be identified and analyzed as a whole; and
c) Record the signals with all the necessary time markers.

The subjects were asked to sit comfortably and keep their left hand still when the experiment started. As a preliminary stage, 30 still, emotionally-neutral pictures were presented to the subject in order to have him/her rest for about 5 minutes. Then the subject played a computer game where he/she repeatedly clicked on the button of a graphic user interface with the label that matched the font color of a word shown on the screen, which named a color. As explained below, the computer game is arranged so that, during selected, identifiable segments of this process, subjects will experience the "Stroop Effect" which has been shown to elicit mental stress in humans.

Considerations for Stress Elicitation

One of the most challenging points in this research is to obtain accurate physiological signals related to mental stress of a human subject. In the context of human-computer interaction, the stress experienced by the user is most likely to be mental (intellectual, emotional and perceptual, as opposed to physical), and moderate in intensity. Physical stressors occur far less frequently (Selye, 1980). Therefore, our experimental protocol seeks to instill moderate mental stress in the participating subjects, at pre-determined times. Accordingly, a computer game based on the well-known 'Stroop Test' was designed and adapted to elicit the mental stress while the subject interacts with the computer.

Computer-based Stress Elicitation

The Stroop Color-Word Interference Test (Stroop, 1935), in its classical version, demands that the color font of a word designating a different color be named. Although there is controversy concerning the exact mechanisms responsible for the Stroop effect, this task has been widely utilized as a psychological or cognitive stressor to induce emotional responses and heightened levels of physiological, (especially autonomic) reactivity (Renaud & Blondin, 1997). We believe that the Stroop test can act reliably as a stress stimulus in the controlled laboratory environment. In our research, a computer-based interacting environment was established to let the subjects experience a similar stress effect. To accomplish this, the classical Stroop Test was adapted into an interactive version that requires the subject to click on the correct answer, which is one of the five buttons shown on the screen, rather than stating it verbally. Since adding task pacing to the Stroop Test might intensify the physiological responses (Renaud & Blondin, 1997), each trial was designed to only wait 3 seconds for a user response. If the subject could not make a decision within 3 seconds, the screen automatically changed to the next trial. One typical example of this trial interface is shown on Figure 1.

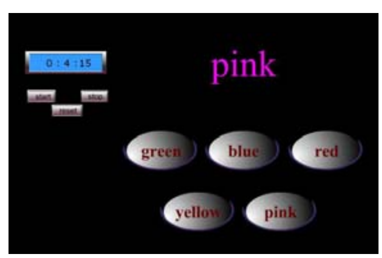

Figure 1. Sample Stroop Test interface

This modified version was implemented with Macromedia Flash® and also programmed to output bursts of a sinusoidal tone through the sound system of the laptop used for stimulation, at selected timing landmarks through the protocol, to time-stamp the physiological signals recorded at those critical instants.

Figure 2 is the audio output schedule in this experiment from the beginning of the game to its end. The complete experiment comprises three consecutive sections. In each section, we have four segments including: 1) 'IS' - the introductory segment to let the subject get used to the game environment, in order to establish an appropriate initial level for the psychological experiment, according to the law of initial values (LIV) (Stern, Ray & Quigley, 2001). 2) 'C' – is a congruent segment, in which the font color and the meaning of the words presented to the user match. There are 45 trials in each congruent segment. The subject should not experience stress in this segment. 3) 'IC' – is an incongruent segment of the Stroop Test in which the font color and the meaning of the words presented differ. This is the segment where a stress response is expected. In total there are 30 trials in this incongruent segment. 4) 'RS' – is a resting segment to let the subject relax for a certain time. The binary number shown in Figure 2 is the de-multiplexed output of the audio signaling used in the system to time-stamp the four physiological signals, BVP, GSR, PD and ST. '01' represents a burst in the left channel audio signal, '10' represents a burst in the right channel and '11' represents simultaneous bursts in both channels. Our previous report on the instrumental setup (Barreto & Zhai, 2003) provides more details on this audio scheme.

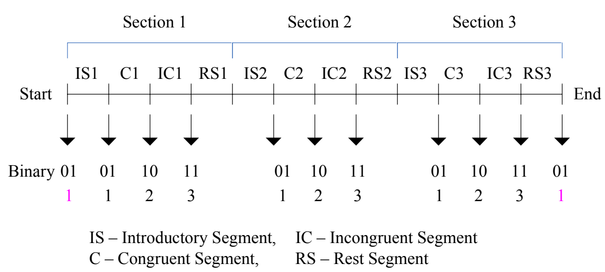

Figure 2. Audio output schedule

Hardware Setup Design

The complete instrumental setup developed for this research is illustrated in Figure 3. The stimulus program (interactive paced Stroop Test) described above runs in a laptop PC. While playing the Stroop Test, the subject has the GSR, BVP and ST sensors attached to his/her left hand. These three signals are digitized, using a multi-channel data acquisition system, NI DAQPad-6020E for USB, a product of National Instrumentation Corp, and the samples are read into Matlab® directly at rate 360 samples/sec. Additionally, the eye gaze tracking system (ASL-504) records PD data to a file on its own interface PC, at rate of 60 samples/sec. The software for this system allows the extraction of selected variables (in this case the pupil diameter and the marker channel) to a smaller file, which in turn can be read into Matlab® also, where it can be aligned with the BVP, GSR and ST signals, thanks to their common timing marks for the start and stop events. At this point the pupil diameter data can be upsampled (interpolated) by six, to achieve a common sampling rate of 360 samples/sec for all four measured signals.

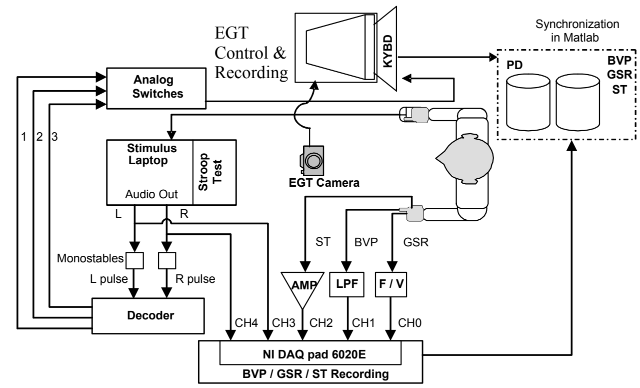

Figure 3. Instrumental Setup

Affective Detection

One key research problem in the Affective Computing area is the mapping between affective states and physiological states (Fernandez & W.Picard, 1998). Using multiple features of the physiological signals to indicate the correlations between a set of raw data and the internal stress sate in each game section is one promising solution to this challenge. In our research a total of 11 features are extracted from each segment of the physiological signals monitored (Figure 4), as described below.

Feature Extraction

The GSR responses in each segment were isolated by using Ktonas' 7-point Lagrangian interpolation algorithm (Ktonas, 1987) on the raw GSR signals. This function is a combination of the 3-point Lagrangian derivative and the 5-point second order data fit algorithm. Let g[n] represent the discrete time sequence obtained by sampling the GSR signal.

$$g''[n] = \frac{2g[n+3] + g[n+2] - 2g[n+1] - 2g[n] - 2g[n-1] + g[n-2] + 2g[n-3]}{20h^2} \qquad (1)$$

Equation (1) can be considered as a 7-point operator, where h is the sampling interval. We multiplied $g''[n]$ by a constant value of 200 to keep the original signal scale. This operator includes the advantages provided by the use of interpolation and data-fit and makes the transient increase in the raw signal 'clearly distinguished from the background'. Using thresholding, individual responses in the GSR signal could be counted and localized. The number of the responses and the mean value in each segment could then be calculated. Also the amplitude and the rising time (from baseline to the peak of response) were recorded as features from each GSR segment.

Additionally, the total area under the rising time curve is treated as the GSR response energy.

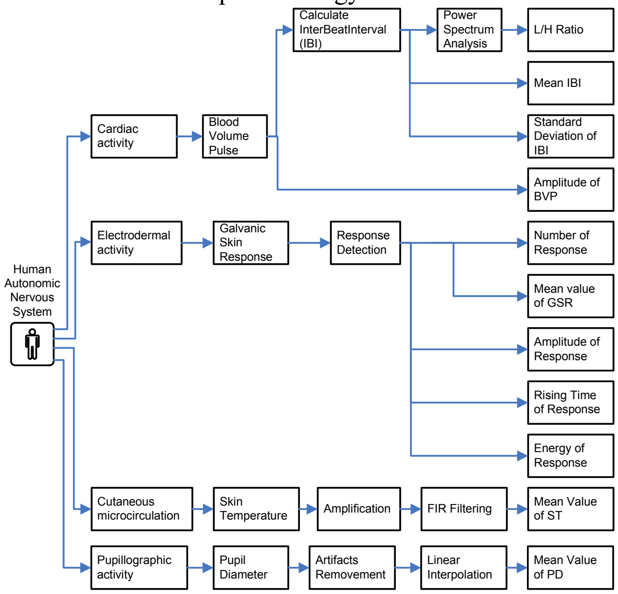

Figure 4. Physiological features extracted

From the BVP signal, each heart beat was first separated and two parameters were generated: P and AM. P is the BVP period, also called interbeat interval (IBI), and defined as the time in milliseconds between two normal, consecutive peaks in the BVP signal. AM is the beat amplitude. The Heart rate Variability (HRV) information, derived from the interbeat intervals calculated, was used to describe the features. Heart Rate variability (HRV) has been extensively applied in understanding the function of the ANS, and has shown a close connection to the emotional state of the subject (Dishman et al., 2000; Rowe, Sibert & Irwin, 1998). The IBI index was analyzed in the frequency domain to overcome the short time duration of the recording. We analyzed the Low Frequency (LF) (0.05-0.15Hz) and High Frequency (HF) (0.16-0.40Hz) bands of the HRV spectrum. The low frequency band reflects sympathetic activity with vagal modulation, and the high frequency band reflects parasympathetic activity. The very low frequency (VLF 0-0.04Hz) domain was not analyzed in this study because VLF assessed from short-term recordings (≤ 5 min) has been shown to be an unreliable measure, according to the Task Force of the European Society of Cardiology and The North American Society of Pacing and Electrophysiology(Electrophysiology, 1996). We expected that increases in the LF/HF ratio could be used to indicate the presence of stress. The mean value and standard deviation of the IBI sequence were also obtained as additional features of the cardiac response. Also, the amplitude of each BVP beat was calculated since its decrease may indicate that the subject is undergoing mental stress (resulting in vasoconstriction).

The amplified skin temperature (ST) signal was firstly filtered to get rid of the noise. The average value of the filtered skin temperature in each segment was then used as a feature element of this signal. It was expected that the temperature in the finger surface would display transient decreases when the stressor stimuli occur.

The raw pupil diameter (PD) signal was recorded separately, as previously described. The artifact gaps due to blinking were filled by interpolation. The feature extracted from the pupil diameter is simply the mean value of PD, which was expected to increase during the stress (incongruent Stroop) segments.

Data Normalization

For each subject, the previous feature extraction techniques were applied to the congruent (C) and incongruent (IC) segments in each of the 3 sections shown in Figure 2, and also to the resting signal recorded when the subjects were watching the still pictures before the experiment started. Let Xs represent the feature value for any of the 11 features extracted from the signals which were recorded during congruent and incongruent segments of the computer game. Let Xr represent the corresponding feature value extracted from the signals which were recorded during the resting period, before the computer game started. To eliminate the initial level due to the individual differences, Equation (2) was first applied to get the corrected feature values for each of the subjects.

$$Ys = \frac{Xs}{Xr} \tag{2}$$

For each subject, there were three C segments and three IC segments. Therefore, six values of any of the features were obtained from the signals recorded during these segments. Equation (3) normalizes each feature value dividing it by the sum of all six segment values.

$$Y_s' = \frac{Ys_i}{\sum_{i=1}^{6} Ys_i} \tag{3}$$

These two stages of normalization proved essential to minimizing the impact of individual subject responses in the training of the learning systems used in our work. After this pre-processing, all features were normalized to the range of [0, 1] using max-min normalization, as shown in Equation (4), to be fed into three learning systems described in the following section.

$$Y_{norm} = \frac{Y'_s - Y'_{s\min}}{Y'_{s\max} - Y'_{s\min}} \tag{4}$$

Stress Recognition

After all the features were extracted, they were provided as input to three types of learning systems, to differentiate the stress states (incongruent Stroop segments) from the normal states (congruent Stroop segments): Naïve Bayes Classifier, Decision Tree Classifier, and Support Vector Machine (SVM).

Naïve Bayes Classifier

The Naive Bayes method is a statistical learning algorithm that applies a simplified version of Baye's rule in order to compute the posterior probability of a category given the input attribute values of an example situation. This classifier is based on probability models that incorporate class conditional independence assumptions (John & Langley, 1995). The method computes the conditional probabilities of the different classes given the values of attributes of an unknown sample and then the classifier will predict that the sample belongs to the class having the highest posterior probability. If an instance is represented by an n-dimensional feature vector, $(x_1,x_2,\ldots,x_n)$, a sample is classified to a class c from a set of possible classes C according to *maximum a posteriori (MAP)* decision rule:

$$\text{classify}(x_1, x_2,..., x_n) = \arg\max_{c\in C} p(C = c)\prod_{i=1}^{n} p(x_i | C = c) \quad (5)$$

The conditional probability in the above formula is obtained from the estimates of the probability mass function using training data.

Decision Tree Classifier

The decision tree classifier is a 'divide-and-conquer' approach. It has a flow-chart-like tree structure, where each internal node tests a particular attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distributions. The basic idea involved is to break up a complex decision into a union of several simpler decisions, hoping the final solution obtained this way would resemble the intended desired solution (Safavian & Landgrebe, 1991). To classify an unknown sample, it is routed down the tree according to the values of the attributes tested in successive nodes, and when a leaf is reached the instance is classified according to the class assigned to the leaf. A path is traced from the root to a leaf node which holds the class prediction for that sample. A J48 decision tree (Witten & Frank, 2005) is used for this classification task.

Support Vector Machines

Support Vector Machines (SVM) are the computational machine learning systems that use a hypothesis space of linear functions in a high dimensional feature space to perform supervised classification (Adjouadi & Ayala, 2004; Adjouadi & Zong, 2005; Joachims, 1997). The support vector machine (SVM) tries to construct a discriminant function for the data points in feature space in such a way that the feature vectors of the training samples are separated into classes, while simultaneously maximizing the distance of the discriminant function from the nearest training set feature vector. SVM classifiers also allow for non-linear discriminant functions by mapping the input vectors into a different feature space using a mapping function Φ: xi $\rightarrow \Phi$(xi), and using the vectors, Φ(xi), xi $\in$ X, as the feature vectors. The corresponding kernel function used by the SVM algorithm is

$$K(x_i, x_k) =< \phi(x_i)\cdot\phi(x_k) > \quad (6)$$

The SVM used in this classification implements John Platt's sequential minimal optimization algorithm (Platt, 1999) for training a support vector classifier.

Performance Measurements

The practical interest of these methods is their capacity to predict the class of previously unseen samples (test set). The original data samples in any given data set are typically divided into a training set and a test set. This is done to have samples available for testing that were never presented to the system during the training phase. Such a strategy for dividing input samples into training and test sets is used in k-fold cross validation techniques (Efron & Tibshirani, 1993). This strategy allows us to train and test on different samples and obviates the need to test on unknown physiological signal samples whose labels (targets) may be uncertain. We used the Weka software, that contains a collection of machine learning algorithms for data mining tasks, for all three learning systems (Witten & Frank, 2005).

Results

Signals from 32 experimental subjects were collected and divided into 192 data entries since each participant generated data under three non-stress (Congruent Stroop) segments and three stress (Incongruent Stroop) segments. Eleven attributes (GSR_{mean}, IBI_{mean}, IBI_{sd}, etc.) were determined for each data entry. After the feature extraction and normalization stages, the data set from each segment had the structure shown in Figure 4. The prediction performance was evaluated using 20-fold cross validation: 20 samples were pulled out as the test samples, and the remaining samples were used to train the classifiers. To evaluate the predictive ability of the classifiers, the total prediction accuracy, which is the number of correctly predicted samples divided by the number of total samples, was calculated for each class.

The goal was to develop and train a system that accepts the various physiological variables as input and predicts the participant's affective state. The Naïve Bayes algorithm, decision tree classifier, and SVM were trained to build the model independently, which could be used to predict the unknown affective state. The overall accuracy reached in each case is listed in Table 1. In our study, the SVM had the highest prediction accuracy of the three approaches.

Table1. **Stress prediction accuracies with three classifiers using physiological features**

Naïve Bayes	Decision Tree	Support Vector Machine
Accuracy (%)	Accuracy (%)	Accuracy (%)
78.65 %	**88.02 %**	**90.10 %**

Conclusions

The results from the research outlined in this paper reinforce the belief that non-invasive, non-intrusive real-time assessment of the affective state of a computer user is likely to become attainable in the future. This paper has shown that, under controlled conditions, the simultaneous monitoring and concurrent processing of four physiological signals: BVP, GSR, ST and PD, yields acceptable levels (up to 90.1%) of differentiation between "relaxed" and "stressed" user states, as elicited by congruent and incongruent Stroop stimulation, respectively. While it is clear that the conditions for the testing reported here are not typical of human-computer interaction (controlled environmental conditions, prolonged blocks of congruent and incongruent stimulation, etc.), this research reveals that, in fact, the affective state of most of our 32 experimental subjects expressed itself to a point where it was detectable by the digital signal processing and machine learning algorithms used.

Acknowledgment

This work was sponsored by NSF grants CNS-0520811, IIS-0308155, HRD-0317692 and CNS-0426125. Ms. Zhai is the recipient of an FIU Dissertation Year Fellowship.

References

Adjouadi, M., Ayala, M. 2004. Introducing Neural Studio: An Artificial Neural Networks Simulator for Educational Purposes. *Computers in Education Journal* 14:33-40

Adjouadi, M., Zong, A. 2005. Multidimensional Pattern Recognition and Classification of White Blood Cells Using Support Vector Machines. *Journal of Particle and Particle Systems Characterization* 22:107-118

Barreto, A., Zhai, J. 2003. Physiological Instrumentation for Real-time Monitoring of Affective State of Computer Users. *WSEAS Transactions on Circuits and Systems* 3:496-501

Dishman, R.K., Nakamura, Y., Garcia, M.E., Thompson, R.W., Dunn, A.L., Blair, S.N. 2000. Heart rate variability, trait anxiety, and perceived stress among physically fit men and women. *International Journal of Psychophysiology* 37:121-133

Efron, B., Tibshirani, R. 1993. An Introduction to the Bootstrap. Chapman and Hall, New York, London

Electrophysiology, T.F.o.t.E.S.o.C.t.N.A.S.o.P. 1996. Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use. *Circulation* 93:1043-1065

Fernandez, R., W.Picard, R. 1998. Signal Processing for Recognition of Human Frustraion. M.I.T

Hudlicka, E. 2003. To feel or not to feel: The role of affect in human-computer interaction. *International Journal of Human-Computer Studies* 59:1-32

Hugdahl, K. 2001. Psychophysiology The Mind-Body Perspective. Harvard University Press

Joachims, T. 1997. Text categorization with support vector machines: learning with many relevant features. *In:* Proc of the10th European Conference on Machine Learning, ECML-98

John, G.H., Langley, P. 1995. Estimating Continuous Distributions in Bayesian Classifiers. *In:* Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. pp. 338-345

Kistler, A., Mariauzouls, C., von Berlepsch, K. 1998. Fingertip temperature as an indicator for sympathetic responses. *International Journal of Psychophysiology* 29:35-41

Ktonas, P.Y. 1987. Automated Spike and Sharp Wave (SSW) Detection. *Methods of Analysis of Brain Electrical and Magnetic Signals, EEG Handbook (revised series)* 1:211-241

Partala, T., Surakka, V. 2003. Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies* 59:185-198

Picard, R.W. 1997. Affective computing. MIT Press, Cambridge, Mass.

Picard, R.W.V., E.; Healey, J. 2001. Toward machine emotional intelligence: analysis of affective physiological state. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23:1175-1191

Platt, J.C. 1999 Fast training of support vector machines using sequential minimal optimization *In:* Advances in kernel methods: support vector learning pp. 185-208 MIT Press

Rani, P., Sarkar, N., Smith, C.A., Adams, J.A. 2003. Affective Communication for Implicit Human-Machine Interaction. *IEEE Transactions on Systems, Man, and Cybernetics.* Under Review

Reeves, B., Nass, C. 1996. The Media Equation. Cambridge Univ. Press

Renaud, P., Blondin, J.-P. 1997. The stress of Stroop performance: physiological and emotional responses to color-word interference, task pacing, and pacing speed. *International Journal of Psychophysiology* 27:87-97

Rowe, D.W., Sibert, J., Irwin, D. 1998. Heart Rate Variability: Indicator of User State as an Aid to Human-Computer Interaction. *CHI*:18-23

Safavian, S.R., Landgrebe, D. 1991. A survey of Decision Tree Classifier Methodology. *IEEE Transactions on Systems, Man, and Cybernetics* 21:660-674

Scheirer, J., Fernandez, R., Klein, J., Picard, R.W. 2002. Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers* 14:93-118

Selye, H. 1980. Selye's Guide to Stress Research. Van Nostrand Reinhold Company

Stern, R.M., Ray, W.J., Quigley, K.S. 2001. Psychophysiological Recording. Oxford University Press

Stroop, J.R. 1935. Interference in serial verbal reactions. *Journal of Experimental Psychology* 18:643-661

Witten, I.H., Frank, E. 2005. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, San Francisco