Concurrent Analysis of Physiologic Variables for the Assessment of the Affective State of a Computer User

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

Departments of Electrical & Computer Engineering[1] and Biomedical Engineering[2]
Florida International University
10555 West Flagler Street, EC-3970
{jazhai002, barretoa}@fiu.edu

Abstract

Much progress has been made during the last 40 years in the quest to improve the interaction of humans with computers. While new modalities of communication between computers and their users continue to be found and enhanced (e.g., speech recognition for human-to-computer communication and speech synthesis for computer-to-human communication), the nature of the exchange between computers and users remains, for the most part, dry and mechanistic. The emerging field of Affective Computing seeks to advance Human-Computer Interaction (HCI) by enabling computers to interact with users in ways appropriate to their affective states. However, a major prerequisite to the fulfillment of the promise of Affective Computing is the development of efficient mechanisms for Affective Sensing, i.e., the ability of the computer to assess the affective state of its user, particularly when it shifts to uncomfortable states, such as stress. Our research pursues the use of three physiological signals: Blood Volume Pulse (BVP), Galvanic Skin Response (GSR) and Pupil Diameter (PD), to automatically monitor the level of stress in computer users. This paper reports on the hardware and software instrumentation development and signal processing approach used to detect the stress level of a subject interacting with a computer, within the framework of a specific experimental task, which is called the 'Stroop Test'. The Stroop Effect is evoked by the mismatch between the font color and the meaning of a certain word (name of a color), displayed to the experimental subject. For this experiment, a computer game was implemented and adapted to make the subject experience this effect, while his/her BVP, GSR and PD signals were continuously recorded. Three data processing techniques were applied to extract effective attributes of the stress level of the subjects throughout the experiment. Current results indicate that there exists a correspondence between changes in those three signals and the shift in the emotional states when stress stimuli are applied to the interaction environment.

1 Introduction

In the last 40 years, computers have been transformed from massive systems that occupied large spaces in "computer rooms" and were programmed through cryptic perforated cards, to miniaturized devices that we carry/wear and are capable to exchange information with us in forms that are readily understandable to most humans (e.g., speech input and output). In spite of this enhancement of the modalities available to users for the exchange of information with the computer, a key characteristic continues to differentiate the fundamental interaction between computers and humans: the mechanistic inflexibility with which most computers carry out their participation in the interaction. Even as the simplest human-human interactions are usually modulated by the mutual appreciation of the "state" of the participants, computers typically will proceed through the interaction, disregarding the "state" of their users. While it may be thought that this is a "subtle" difference in the quality of the interaction, Reeves and Nass, from Stanford University, have long postulated that human-computer interaction is inherently natural and social, following the basics of human-human interaction (Reeves & Nass, 1996). This level of interaction between humans and computers will be needed to truly fulfil the vision of men and machines engaged as genuine collaborative systems, integrated human–machine systems, and joint cognitive systems (Hollnagel, 2003). Affective Computing, introduced by Picard as 'Computing which relates to, arises from, or deliberately influences emotions' (Picard, 1997), seeks to endow computers with the capabilities that would support that kind of interaction. However, a major challenge that still stands in the way of achieving that goal is the difficulty of having computers assess the affective state of their users. Our research attempts to visualize and evaluate the emotional state identified with 'stress' of the computer users, through several physiological signals that can be measured non-invasively and non-intrusively, and its preliminary results may be projected to support the emerging field of Affective Computing. The previous research in our group has developed the hardware and software integration setup to collect the users' affective (especially stress) signals. In the research reported in this paper, we improve the instrumentation setup to obtain more reliable and stable signals and three machine learning technologies are implemented on the collected signals. The results are compared in order to identify the most successful method used in the realization of stress detection.

2 Detection of stress through measurement of physiological variables

Most of us can attest to some clear, involuntary and unmaskable changes in our bodies as reactions to strong emotional stimuli: our hearts may change their pace during climatic moments in a sports event we witness; our hands may turn cold and sweaty when we are scared; we may feel "a rush of blood to the head", when we get into a strong argument. These are not imaginary changes, but instead reflect the perception of an actual reconfiguration of our organism that takes place as a reaction to the psychological stimuli listed.

Just like we are capable of identifying an affective shift in another human by sensing his/her physiological reconfiguration (e.g., seeing the redness in the face of an angry colleague, feeling the wetness and cold of a fearful person's hand) computers could, potentially, measure these physical quantities from their users and utilize those measurements to assess their affective states.

2.1 Physiological aspects of sympathetic activation

Physiological signals can be treated as a truly reliable indicator of the emotional state since they are not voluntarily manipulated ('faked') by the computer users. From human physiology studies, it is well known that the sympathetic division of the human Autonomic Nervous System (ANS) significantly influences some physiological variables. The heart rate (HR), blood pressure (BP), Galvanic skin resistance (GSR), blood volume pulse (BVP) and pupil diameter (PD) are all affected by branches of the sympathetic division of the ANS. We collected three physiological variables (GSR, BVP, PD) simultaneously to analyze potential concurrent changes that may be due to sympathetic activation associated with 'stress'. We selected these three signals based on the fact that they can be conveniently monitored by non-invasive means.

3 Experimental Methodology

3.1 Instrumental setup

Six healthy male subjects, aged from 21 to 35 years old, participated in this study. They were recruited from the student body of our FIU College of Engineering. To assure reliability of the changes in PD, the lighting of the environment was kept at a constant level during the experiment and the illumination of the system was the same for all the subjects. A GSR2 module, by Thought Technology LTD (West Chazy, New York) was used to measure the Galvanic Skin Response (GSR) and photoplethysmography (PPG) was used to measure the blood volume in the skin capillary bed, of the left ring finger. Figure 1 shows the GSR and BVP sensors used. The sampling rate for GSR and BVP 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. Figure 2 shows the Eye Gaze Tracking system used. The detailed description of these sensors and the signal collection and associated synchronization method can be found in our previous report (Barreto & Zhai, 2003). The stimulus program (interactive Stroop Test), described in Section 3.2, ran in a laptop PC. While undergoing the Stroop Test, the subject had the GSR and BVP sensors attached to his left hand. Both GSR and BVP signals were converted, using a multi-channel data acquisition system, NI DAQPad-6020E for USB (National Instrumentation Corp) directly at a rate of 360 samples/sec into a file readable by Matlab®. Additionally, the eye gaze tracking system (ASL-504) had been calibrated and recorded PD data to a file on its own interface PC, at a 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 and GSR 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 three measured signals. The complete instrumental setup used is illustrated in Figure 3.

Figure 1: Galvanic Skin Resistance (GSR) sensor (white) and Photoplethysmography (PPG) sensor (black) used to measure the subject's Blood Volume Pulse (BVP) signal, shown as attached to the subject during the experiments.

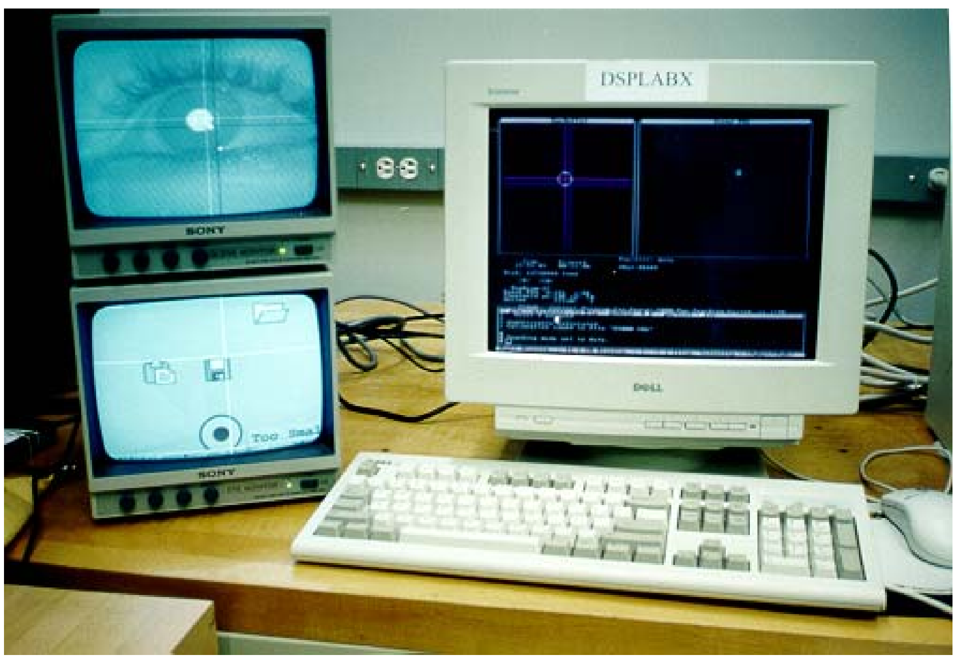

Figure 2: Eye Gaze Tracking (EGT) system used to record the pupil diameter (PD) signal.

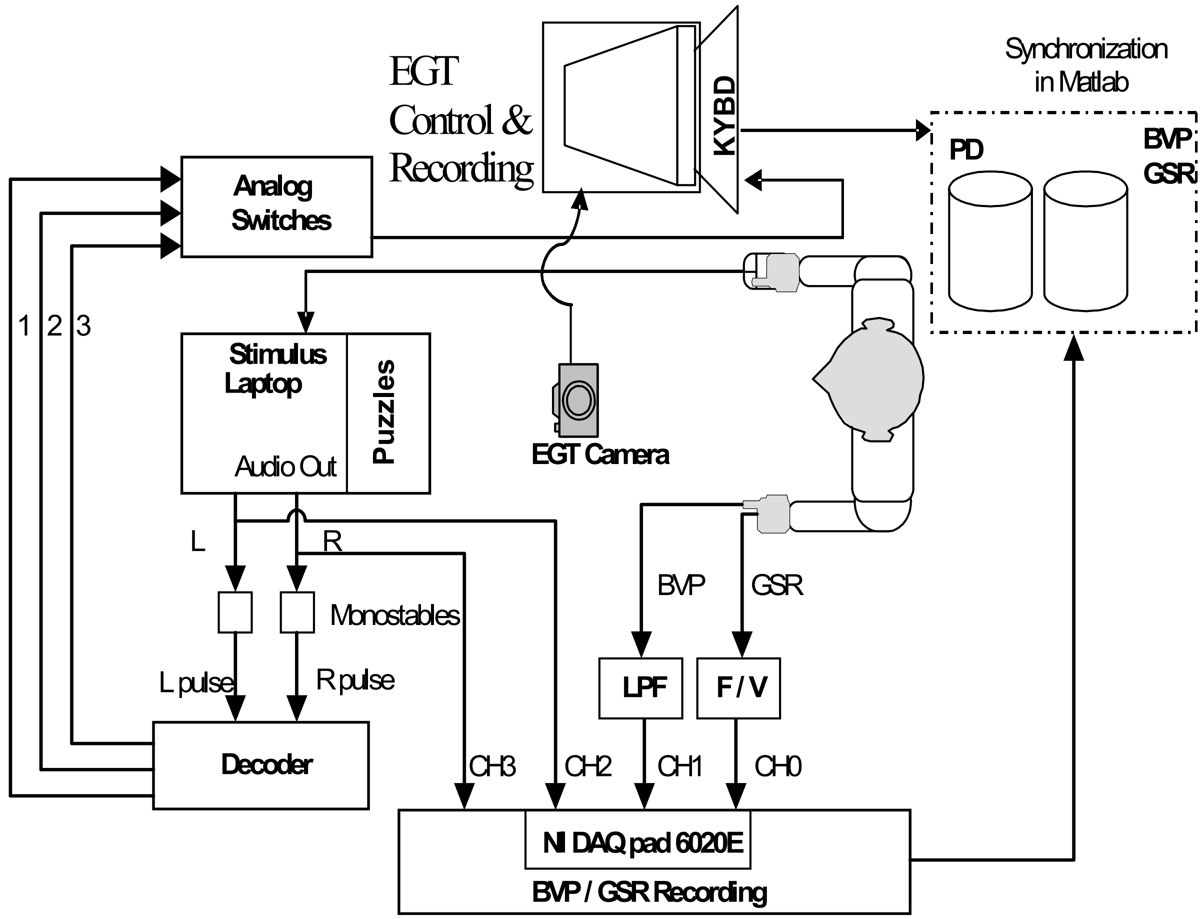

Figure 3: Instrumentation setup.

3.2 Elicitation of affective shifts

One of the most challenging points in this research is to provide adequate stimulation to reliably elicit mental stress in a human subject. Design and implementation of experiments simulating mental stress require considerable insight into human psychology. In the context of human-computer interaction, the stress experienced by the user is most likely to be mental (as opposed to physical), and moderate in intensity. Therefore, our experimental protocol sought to produce 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 was interacting with the computer.

In the classical version of the Stroop Color-Word Interference Test (Stroop, 1935) the subject is required to name the font color of a word designating a different color. 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). For our research, an interacting environment needed to be established to let the subjects experience a similar effect. To accomplish this, the classical Stroop Test was adapted into an interactive version that requires the subject to click the correct answer, which is one of the five buttons shown on the screen, rather than stating it verbally. One typical example of this test interface is shown on Figure 4.

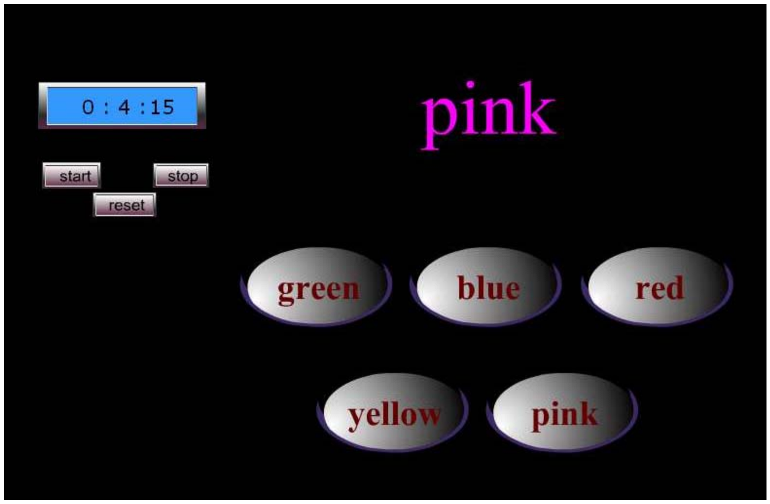

Figure 4: Sample Stroop Test interface.

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 recorded signals at those critical instants. Our previous report on the instrumental setup (Barreto & Zhai, 2003) provides more details on the audio schemes to achieve the desired time-stamping in the three recorded signals. Figure 5 is the audio output schedule in this experiment from the beginning of the game to its end. The complete experiment comprises three consecutive sequences. In each sequence, there are the following segments:

- ‘IS’ - An introductory section to let the subject get used to the game environment, in order to establish an appropriate initial level in the psychological experiment, according to the law of initial values (LIV) (Stern, Ray & Quigley, 2001).
- ‘C’ - A congruent segment, in which the font color and the meaning of the words presented to the user match.
- ‘IC’ - An incongruent segment of the Stroop Test in which the font color and the meaning of the words presented differ.
- ‘RS’ - A resting section 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 three physiological signals, BVP, GSR and PD. ‘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.

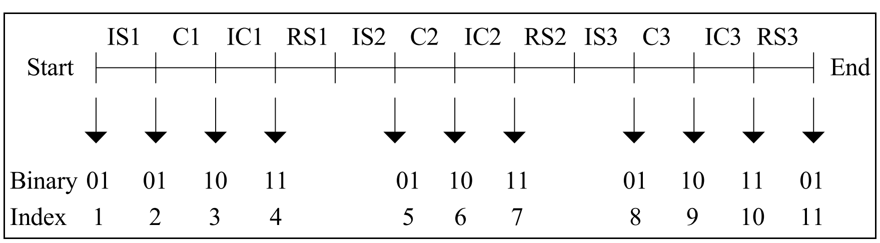

Figure 5: Audio output schedule.

4 Signal Analysis

4.1 Extraction of relevant features

In Affective Computing, one key research problem is the mapping between affective states and physiological states. This is a question that has been investigated in the psychophysiology community for a long time (Fernandez & Picard, 1998). To indicate the correlations between a set of raw data and the internal stress sate in each test section, a set of significant feature signals were extracted from each physiological measure. Multiple parameters were extracted from each physiological measure, as shown on Figure 6.

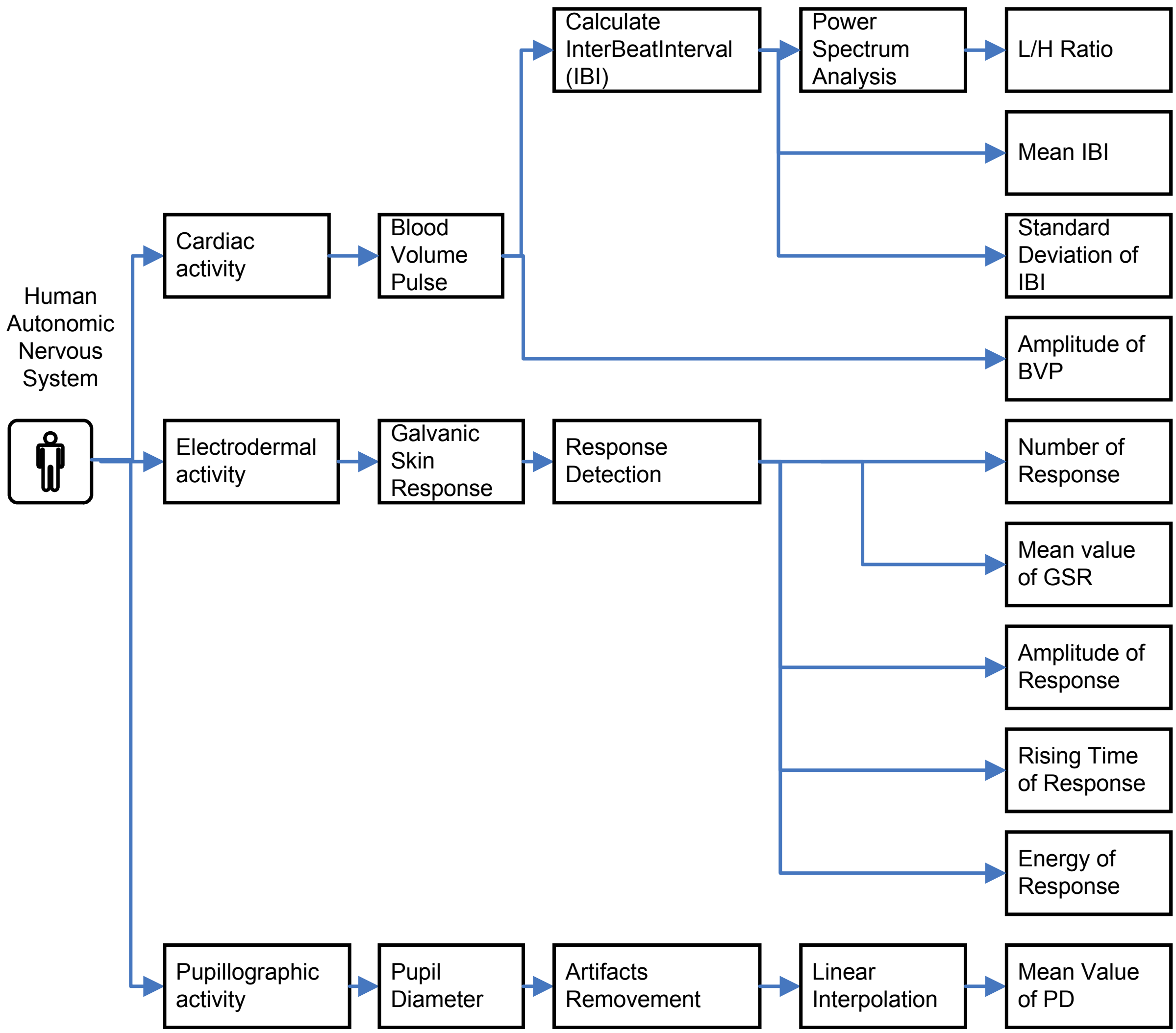

Figure 6: Physiological features extracted.

For the GSR signal, all the responses were isolated by using Ktonas' 7-point Lagrangian interpolation algorithm (Ktonas, 1987) on the raw GSR signals. In his paper, he proposed the measurement of the second time derivative with 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.

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

Equation (1) can be considered as a 7-point operator. In our processing, $g''[n]$ was multiplied 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'. By thresholding, individual responses in the GSR signal could be counted and localized. The number of the responses and the mean value in each section 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 section. Additionally, the total area under the rising time curve was treated as the GSR response energy.

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), 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 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). Clinically, HRV has been obtained through the electrocardiogram (EKG) signal, but in the context of human-computer interaction studies, photoplethysmographic (PPG) measurement of the Blood Volume Pulse (BVP) was preferred as a much more convenient monitoring method. The IBI index was analyzed in the frequency domain to overcome the short time duration of the recordings. We analyzed the Low Frequency (LF) (0.05-0.15Hz) and High Frequency (HF) (0.16-0.40Hz) bands of the HRV. 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 ("Heart Rate Variability", 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.

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 was a simple parameter: the mean value of PD. According to the introduction, we expected that the mean PD should increase during the stress (incongruent Stroop) segments.

4.2 Approaches used for the recognition of stress

After all the features were generated, they were provided as input to learning systems, to differentiate the stress state (incongruent Stroop segments) from the normal state (congruent Stroop segments) of a person working on a computer task. In particular, three popular learning algorithms were employed for this learning and classification process: Naïve Bayes, Decision Tree Classifier, and Support Vector Machine (SVM).

4.2.1 Naïve Bayes Classifier

Naïve Bayes classifiers are based on probability models that incorporate class conditional independence assumptions (John & Langley, 1995). This 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 a posteriori 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:

$$\text{classify}(x_1, x_2, \ldots, x_n) = \arg\max_{c \in C} p(C = c) \prod_{i=1}^{n} p(x_i | C = c) \quad (2)$$

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

4.2.2 *Decision Tree Classifier*

The decision tree classifier is a flow-chart-like tree structure, where each internal node denotes a test on an 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). In order to classify an unknown sample, the attribute values of the sample are tested against the decision tree. In contrast to conventional single-stage classifiers where each data sample is tested against all classes, in a tree classifier a sample is tested against only certain subsets of classes, thus eliminating unnecessary computations. A path is traced from the root to a leaf node, which holds the class prediction for that sample.

4.2.3 *Support Vector Machines*

Support Vector Machines (SVMs) are the computational machine learning systems that use a hypothesis space of linear functions in a high dimensional feature space to perform supervised classification (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 $\Phi: x_i \rightarrow \Phi(x_i)$, and using the vectors, $\Phi(x_i)$, $x_i \in X$, as the feature vectors. The corresponding kernel function used by the SVM algorithm is:

$$K(x_i, x_k) = \langle \Phi(x_i).\Phi(x_k) \rangle \quad (3)$$

Standard kernel functions include polynomial, radial basis function (RBF) and sigmoid kernel functions.

4.2.4 *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, which contains a collection of machine learning algorithms for data mining tasks, for naïve Bayes and decision tree classifiers (Witten & Frank, 2000), and the LibSVM software package for SVM (Chang & Lin, 2001).

Signals from six experimental subjects have been collected and divided into 36 data entries. Each participant generated data under three non-stress (Congruent Stroop) segments and three stress (Incongruent Stroop) segments. Ten attributes (GSR_{mean}, IBI_{mean}, IBI_{sd} , etc.) were determined for each data entry. After the feature extraction, the data set had the structure shown in Figure 6.

The prediction performance was evaluated using the jackknife test (Efron & Tibshirani, 1993); each sample was singled out in turn as a test sample, 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.

5 Results

The goal was to develop and train a system that accepts the various physiological variables as input and identifies the participant's affective state. The naïve Bayes, decision tree classifier, and SVM were trained to build the model independently, which could be used to identify the unknown affective state. For the SVM, three standard kernel functions were tried on this affective data set. The overall accuracy reached in each case is listed in Table 1. Amongst the three kernels used with SVM, the sigmod kernel gave the best prediction performance, 80%, and the other two kernels showed similar performance, 57.14% and 60%, respectively, while the prediction accuracies of the naïve Bayes and the decision tree classifiers were 58.33% and 69.44% respectively. The SVM classifier with sigmoid kernel achieved the highest prediction accuracy over all five approaches.

Table 1: Stress prediction accuracies with Naïve Bayes, Decision Tree and SVM classifiers using physiological features

Naïve Bayes	Decision Tree	SVM Classifiers		
		Linear Kernel	RBF Kernel	Sigmoid Kernel
Accuracy (%)	Accuracy (%)	Accuracy (%)	Accuracy (%)	Accuracy (%)
58.33%	**69.44%**	**57.14%**	**60%**	**80%**

6 Discussion and Conclusions

The results indicating the classification accuracy of the stress episodes show a promising correlation between the emotional stress and the physiological signals monitored. Additionally, the classifier developed in this project gives more flexibility in defining the feature vectors used for classification. The simplification of the BVP, GSR and PD data into the features derived from them has proven advantageous in providing only essential information to the classification system.

From the tests performed with SVM classifiers in this project, the sigmoid kernel function proved more suitable for this physiological application.

Although the SVM proposed in this work was able to reach 80% accuracy in differentiating the stress state from the normal working state, other classification methods should still be explored, and their potential to provide an even higher accuracy in the identification of stress episodes evaluated. Some alternative approaches that may be successful in this classification task include: dynamic neural networks, hidden Markov model, etc. One other direction is to ensemble models from different learning algorithms effectively and use the combined model to predict new samples.

Additionally, larger collections of experimental data should be gathered, to allow for the development of a stronger classifier.

7 Acknowledgements

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

References

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

Chang, C.-C. and Lin, C.-J. "LIBSVM: a Library for Support Vector Machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm," 2001.

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

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

Fernandez, R. and Picard, R. W., "Signal Processing for Recognition of Human Frustraion," M.I.T 447, 1998.

Hollnagel, E., From human factors to cognitive systems engineering: human machine interactions in the 21st century. Tokyo: ERC Publishing, 2003.

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

John, G. H. and Langley, P., "Estimating Continuous Distributions in Bayesian Classifiers," presented at Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, 1995.

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

Picard, R. W., *Affective Computing*. Cambridge,MA: M.I.T. Press, 1997.

Reeves, B., and Nass, C., The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. New York, Cambridge University Press, 1996.

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

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

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

Stern, R.M., Ray, W.J., and Quigley, K.S., “Psychophysiological Recording”, 2nd Edition, Oxford University Press, New York, NY, 2001.

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

Task Force of the European Society of Cardiology and The North American Society of Pacing and Electrophysiology. Electrophysiology, "Heart Rate Variability : Standards of Measurement, Physiological Interpretation, and Clinical Use," *Circulation*, vol. 93, pp. 1043-1065, 1996.

Witten, I. H. and Frank, E., *Data Mining: Practical machine learning tools with Java implementations*: Morgan Kaufmann, 2000.