# USER STRESS DETECTION IN HUMAN-COMPUTER INTERACTIONS

**JING ZHAI[1], ARMANDO B. BARRETO [1,2], CRAIG CHIN[1], CHAO LI[1]**
[1]Electrical and Computer Engineering Department, [2]Biomedical Engineering Department
Florida International University

## KEYWORDS

Blood Volume Pulse, Galvanic Skin Resistance, Pupil Diameter, Stroop Test, Stress Detection, Affective Computing, Human-Computer Interaction

## ABSTRACT

The emerging research area of Affective Computing seeks to advance the field of Human-Computer Interaction (HCI) by enabling computers to interact with users in ways appropriate to their affective states. Affect recognition, including the use of psychophysiologcal measures (e.g. heart rate), facial expressions, speech recognition etc. to derive an assessment of user affective state based on factors from the current task context, is an important foundation required for the development of Affective Computing. Our research focuses on 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'. For this experiment, a computer game was implemented and adapted to make the subject experience the Stroop Effect, evoked by the mismatch between the font color and the meaning of a certain word (name of a color) displayed, while his/her BVP, GSR and PD signals were continuously recorded. Several 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 interesting similarity among changes in those three signals and the shift in the emotional states when stress stimuli are applied to the interaction environment.

## INTRODUCTION

In the future, computers will be expected to communicate with humans in a variety of applications. In some research communities we no longer even speak of users and machines as separate entities, but rather of collaborative systems, integrated human–machine systems, and joint cognitive systems[1]. Affective Computing, a relatively new area of computing research, has been described as 'Computing which relates to, arises from, or deliberately influences emotions[2].' 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 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.

## METHODS

### 1. Stress elicitation

In the context of human-computer interaction, the stress experienced by users is most likely to be mental (as opposed to physical), and moderate in intensity. Therefore, to mimic the real world stress environment, the experimental protocol running in the laboratory environment should be carefully set up. The Stroop Color-Word Interference Test[3], in its classical version, demands that the color of a word designating a different color be named. In our research, 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. 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. The complete experiment comprises three consecutive sequences. In each sequence, we have the Introductory Section (IS), Congruent segment (C), Incongruent segment (IC) and Resting Section (RS). A more detailed description of the overall experimental setup has been reported previously by our group[4].

## 2. Physiological aspects of stress detection

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 setup

Six healthy male subjects participated in this study, aged from 21 to 35 years old, recruited from the student body of our FIU College of Engineering. To assure reliability of the changes of 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 collect 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 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. The detailed description of these sensors and the signal collection and associated synchronization method can be found in our previous report[4]. The stimulus program (interactive Stroop Test) described above ran in a laptop PC. While playing 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 1.

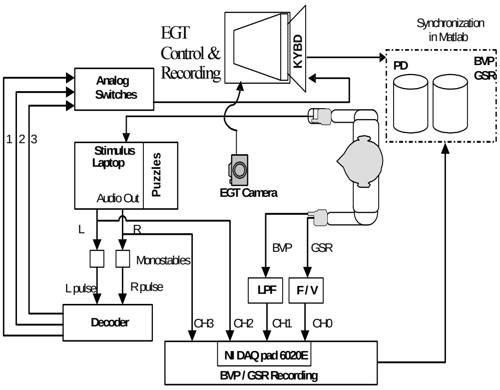


Figure 1: Instrumentation setup

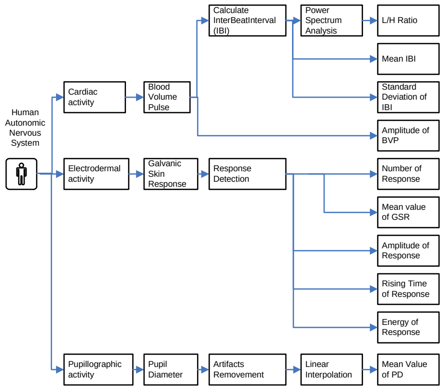


Figure2. Physiological features extracted

## 4. Feature extraction

In the Affective Computing area, one key research problem is the mapping between affective states and physiological states. This is a question that is still being investigated in the psychophysiology community[5]. To indicate the correlations between a set of raw data and the internal stress sate in each game section, a set of significant feature signals must be extracted from each physiological measure. Multiple parameters were extracted from each physiological measure, as shown on Figure2. For the GSR signal, all the responses were isolated by using Ktonas' 7-point Lagrangian interpolation algorithm[6] 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 sequenceobtained 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. We multiplied $g''[n]$ by 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 is 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 ANS, and has shown a close connection to the emotional state of the subject[7, 8]. 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 recording. 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[9]. We expect 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 have to be filled by interpolation. The feature extracted from the pupil diameter is a simple parameter: the mean value of PD. According to the introduction, we expect that the mean PD should increase during the stress (incongruent Stroop) segments.

## 5. Stress Recognition

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: Naïve Bayes, Decision Tree Classifier, and Support Vector Machine (SVM) for this learning and classification process.

### *5.1 Naïve Bayes Classifier*

Naïve Bayes classifiers are based on probability models that incorporate class conditional independence assumptions [10]. 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 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: $\mathrm{classify}(x_1,x_2,\ldots,x_n)=\arg\max_{c\in C} p(C=c)\prod_{i=1}^{n} p(x_i|C=c)$. The conditional probability in the above formula is obtained from the estimates of the probability mass function using training data.

### *5.2 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 [11]. 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.

### *5.3 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 [12]. 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 $K(x_i, x_k) = <\Phi(x_i).\Phi(x_k)>$. Standard kernel functions include polynomial, radial basis function (RBF) and sigmoid kernel functions.

### *5.4 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 [13]. 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 Naïve Bayes and decision tree classifiers [14], and the LibSVM software package for SVM [15].

## RESULTS

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 2.

The prediction performance was evaluated using the jackknife test [13]; 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.

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, decision tree classifier, and SVM were trained to build the model independently, which could be used to predict 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. For three kernels in SVM, the sigmod kernel gives the best prediction performance, 80%, and the other two kernels show similar performance, 57.14% and 60% respectively, while the prediction accuracy of Naïve Bayes and decision tree is 58.33% and 69.44% respectively. The SVM with sigmoid kernel has 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%** |

## DISCUSSION & 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 form 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.

## ACKNOWLEDGMENTS

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

## REFERENCES

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

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

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

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

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

[6] P. Y. Ktonas, "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.

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

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

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

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

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

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

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

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

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