Fifth LACCEI International Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2007)
"Developing Entrepreneurial Engineers for the Sustainable Growth of Latin America and the Caribbean:
Education, Innovation, Technology and Practice"
29 May – 1 June 2007, Tampico, México.

Potential of Pupil Diameter Monitoring for the Detection of Affective Changes in Human-Computer Interaction

Armando Barreto
Florida International University, Miami, Florida, USA, barretoa@fiu.edu

Naphtali Rishe
Florida International University, Miami, Florida, USA, rishen@cis.fiu.edu

Jing Zhai
Florida International University, Miami, Florida, USA, jzhai002@fiu.edu

Ying Gao
Florida International University, Miami, Florida, USA, ygao002@fiu.edu

ABSTRACT

There is continued interest in mechanisms that could inform computers of the affective state of their users. Some proposed approaches analyze physiological signals from the users to asses their affective states. Most of these solutions focus on physiological signals traditionally used for "detection of deception" (lie detector) experiments, such as the Galvanic Skin Response (GSR), the Blood Volume Pulse (BVP), or the Skin Temperature (ST). We monitored these signals, as well as the Pupil Diameter (PD), during a sequence of congruent and incongruent Stroop trials, expected to result in intervals of relaxation and stress in the subject (respectively). Single feature signals were derived from these variables (GSR mean, BVP interbeat interval mean, ST slope and PD mean) and used as detection signals for the identification of incongruent Stroop segments (i.e., stressed states in the subject). Receiver Operating Characteristic (ROC) analyses were carried out for each of the four single detection signals, as a mechanism to compare their discriminating power. The PD signal was found to result in the higher area under the ROC curve, which may imply a stronger potential for the discrimination of the stressed segments elicited in the subject by incongruent Stroop stimulation.

Keywords: Affective Sensing, Pupil Diameter, Human-Computer Interaction, Receiver Operating Characteristic

1. INTRODUCTION

As computers pervade virtually every aspect of our activities in society, playing critical roles in the ways we communicate, work, learn and socialize, the need for more natural and less mechanistic interfaces between machines and their users is becoming a central concern for the evolution of computer systems. As highlighted early on by Picard (Picard, 1997), one of the key aspects that are missing in our interaction with current-day computers is their complete lack of responsivity to the affective states of their users. While this machine disregard for the emotional state of the user may be of secondary importance for some computer applications, e.g., an Automated Teller Machine (ATM), it may be critical for the fulfillment of the main purpose of some computer-based systems. For example, it has been clearly established that Intelligent Tutoring Systems (ITS) would benefit from recognizing the emotional state of the learner, as there are emotional conditions that will lead to more efficient learning (Chaffar & Frasson, 2005)(Chaffar & Frasson, 2006).

However, one of the major impediments in implementing affective computing systems is the difficulty to obtain reliable real-time assessments of the affective states experienced by the user. The assessment of affective states is, in itself, a challenging task, which is only further complicated by the need to perform it without overt cooperation from the subject in the context of human-computer interaction (as the affective assessment should not distract the user from his/her primary task in the interaction, e.g., learning). Working within these limitations, a variety of methods for measuring affective states in the users have been tried, such as the identification of facial expressions, in isolation, or in combination with speech understanding and body gesture recognition (Cowie et al., 2001). Another important subgroup of efforts have been directed to the evaluation of the affective state of the user through analysis of some of his/her physiological signals, which can be monitored non-invasively. Since physiological variables in humans are inherently controlled by their autonomic nervous system, these expressions of emotion are less susceptible to environmental interference or voluntary masking than others, such as, for example, facial expression or speech activity.

2. RATIONALE FOR PHYSIOLOGICAL MONITORING OF USER AFFECTIVE STATES

The monitoring of physiological changes as a window into the realm of affective states in a subject is grounded on well established associations between the divisions of the Autonomic Nervous System (ANS) and numerous physiological processes around the body. The Autonomic Nervous System (ANS) coordinates the cardiovascular, respiratory, digestive, urinary and reproductive functions according to the interaction between a human being and his/her environment, without instructions or interference from the conscious mind (Martini et al., 2001). The ANS is studied as composed of two divisions: The Sympathetic Division and the Parasympathetic Division. The Parasympathetic Division stimulates visceral activity and promotes a state of "rest and repose" in the organism, conserving energy and fostering sedentary "housekeeping" activities, such as digestion (Martini et al., 2001). In contrast, the Sympathetic Division prepares the body for heightened levels of somatic activity that may be necessary to implement a reaction to stimuli that disrupt the "rest and repose" of the organism. When fully activated, this division produces a "flight or fight" response, which readies the body for a crisis that may require sudden, intense physical activity. An increase in sympathetic activity generally stimulates tissue metabolism, increases alertness, and, globally, helps the body transform into a new status, which will be better able to cope with a state of crisis. Parts of that re-design or transformation may become apparent to the subject and may be associated with measurable changes in physiological variables, which leads to the possibility of detecting and studying changes in autonomic activity, and particularly increases in sympathetic activation.

We have investigated the changes in a reduced set of physiological signals influenced by the activity of the Autonomic Nervous System, in computer users faced with a task known to elicit moderate levels of stress, in an attempt to use the analysis of those signals for identification of the presence of stress in the subject. Three of the signals investigated are frequently monitored in psychophysiological studies and in "detection of deception" (lie detector) instrumentation. These are the Galvanic Skin Response (GSR), the Blood Volume Pulse (BVP) signal and the measurement of Skin Temperature (ST) in the extremities (in our case in the thumb). Our study also included continuous measurement of the Pupil Diameter (PD) of the left eye of the subjects as they performed the computer task. Our initial approach was to obtain several features from these four signals and apply machine learning algorithms to classify these groups of features towards the detection of those instances that corresponded to segments of the records associated with the presence of stress in the subject. Previous reports have disseminated the results we have obtained through that synergistic approach, which involved all four monitored signals (Zhai & Barreto, 2006a) (Zhai & Barreto, 2006b) (Barreto & Zhai, 2004). However, as the data analysis progressed, we have been attracted by the fact that the Pupil Diameter (PD) signal seemed to be particularly powerful in discriminating the intervals of our experiments in which stimuli had been applied to elicit stress in the

subject. In this paper we report on the comparison we have carried out between the pupil diameter and features obtained from the three other physiological signals monitored in our experiments. The Receiver Operating Characteristic (ROC) analysis performed shows that the pupil diameter signal has a strong potential for use as a key indicator of stress in computer users. We believe that this is an important observation, as pupil diameter measurements are not included in most contemporary physiological measurement systems used for affective assessment.

3. STRESS ELICITATION PROTOCOL

In order to elicit mental stress at controlled intervals in 32 experimental subjects, a computerized "Paced Stroop Test" was used. The classical Stroop Test was adapted into an interactive version requiring the subject to click on the on-screen button identifying the font color of a word displayed (regardless of the meaning of the word shown). It has been proposed (Renaud & Blondin, 1997) that adding task pacing to the Stroop Test might intensify the physiological responses. Therefore, 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. This stimulus paradigm was implemented with Macromedia Flash®. The Flash program also output bursts of sinusoidal tones 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 audio output schedule for the experiment from the beginning of the session to its end is shown in Figure 1.

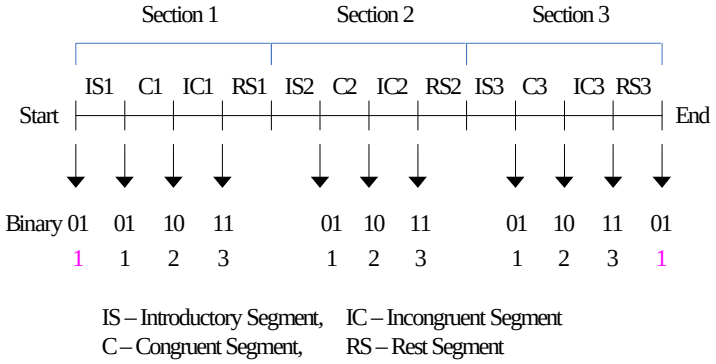

Figure 1: Experiment sequence

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 task environment, in order to establish an appropriate initial level for his/her psychological state, according to the law of initial values (LIV) (Stern et al., 2001); 2) 'C' – is a Congruent segment, comprising 45 Stroop congruent word presentations (font color matches the meaning of the word), which are not expected to elicit significant stress in the subject; 3) 'IC' – is an Incongruent segment of the Stroop Test in which the font color and the meaning of the 30 words presented differ, which is expected to induce stress in the subject; 4) 'RS' – is a Resting Segment to let the subject relax for some time. The binary numbers shown in Figure 1 represent the stereo audio code used in the system to time-stamp the four physiological signals, BVP, GSR, PD and ST. Our previous report on the instrumental setup (Barreto & Zhai, 2004) provides more details on this audio scheme.

4. PHYSIOLOGICAL MONITORING

While performing the Stroop Test, the subject had the GSR, BVP and ST sensors attached to his/her left hand. These three signals were digitized, using a multi-channel data acquisition system, NI DAQPad-6020E, and the samples were read into Matlab® directly at a rate of 360 samples/sec. The pupil diameter measurements were obtained through an infrared eye gaze tracking system (ASL-504), at a rate of 60 samples/sec (Figure 2). The environmental illumination in the room where the tests took place was kept constant during the experiment and uniform across subjects.

Figure 2: . Infrared Camera of the Eye Gaze Tracking System used to measure the pupil diameter

The hardware setup used for the experiment provided for the insertion of event marks in the PD record at the same points in the experiment as indicated by the binary codes in Figure 1. The resulting file was later exported to Matlab® , where the PD signal was upsampled to a sampling rate of 360 samples/second and aligned with the BVP, GSR and ST signals, according to their common timing marks for the start and stop events.

Figure 3 shows an example of the four signals recorded from a subject through the complete length of the experimental session, after all of the signals have been synchronized (at a sampling rate of 360 samples per second). The gaps in the pupil diameter signals, due to blinking, have been compensated by automatic interpolation. Figure 4 shows an example of the complete PD signal. The beginning of each of the experimental segments has been indicated with a vertical line and labeled with a single letter to the left of the corresponding line: C = Congruent Stroop segment; I = Incongruent Stroop segment and R = Rest segment.

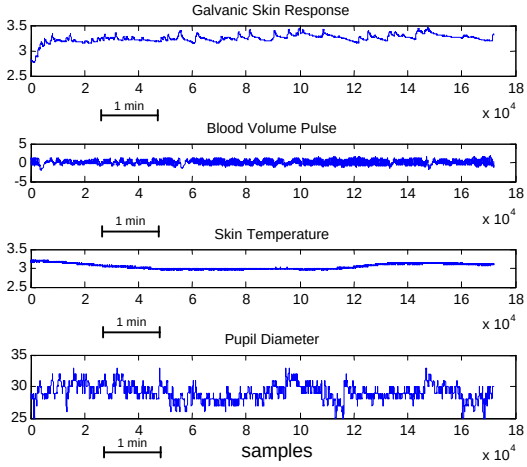

Figure 3: . GSR, BVP, ST and PD signals (top to bottom) recorded during a complete experimental session

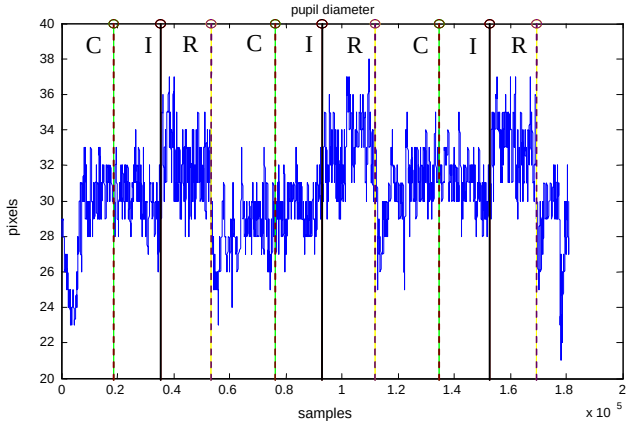

Figure 4: . Example of a PD signal with segment boundary lines superimposed

It should be noted in Figure 4 that an increase in PD signal level can be observed at each start of the incongruent Stroop segments (solid vertical lines labeled "I" in the figure), in spite of the low resolution available for the PD measurements. This increase in PD is reverted on the transition (R) from each incongruent Stroop segment to the

following rest segment. It should also be noticed that none of the other signals in Figure 3 seems to display such close correlation with the sequence of Stroop segments.

Signals from 32 experimental subjects were collected and divided into 192 data entries, since each participant generated data under three relaxed (congruent Stroop) segments and three stressed (incongruent Stroop) segments.

5. EXTRACTION OF NORMALIZED DETECTION SIGNALS

In order to compare the potential of single-index indicators derived from the four physiological signals measured towards the identification of the congruent and incongruent Stroop segments, normalized features were derived from each of the four signals. The sample values of each of the signals were consolidated in a single feature value for each congruent or incongruent segment in the test, following the procedure described below.

The average value of the GSR samples collected during the whole extent of a congruent or incongruent Stroop segment was used as a representative response for this variable for each segment: GSRmean. Increased sweat production during "stressed" segments would predict a noticeable change of this average value during those segments. The interbeat interval (IBI) sequence was determined as the time differences between peaks in the BVP signal, for every two consecutive beats. The inverse of the IBI, expressed in beats per minute (BPM) is the heart rate, which is known to be altered by autonomic activation. The average of the IBI values in each segment, the BVPIBImean value was extracted as a representative feature for this signal. For the skin temperature signal, it was expected that the temperature in the finger surface would display transient decreases when the stressor stimuli occur. Therefore, the amplified ST signal was first filtered to remove recording noise and then a digital low-pass differentiation algorithm was used to obtain the slope of the ST signal, STslope. The pupil diameter (PD) signal was first pre-processed to handle gaps due to blinking. The blinking gaps were automatically detected and filled by interpolation. The single-index signal extracted from the pupil diameter samples in each segment was simply the average value of PD (which we have labeled simply "PD"). As observed in Figure 4, the PD values increased during periods of stress (incongruent Stroop segments), as expected according to previous independent reports from the psychophysiology research literature (Siegle et al., 2004) (Steinhauer et al., 2004).

In order to compare the discriminant power of these single-index signals within the same framework they all were normalized. Let Xs represent the feature value for any of the raw features defined from the signal sample values during congruent and incongruent segments of the experiment. Let Xr represent the corresponding feature value extracted from the signals samples that were recorded during the relaxation period, prior to the first congruent Stroop segment. To eliminate the initial level due to the individual differences, Equation (1) was first applied to get the corrected feature signals (Ys) for each of the subjects.

$$Y_s = \frac{X_s}{X_r} \tag{1}$$

For each subject, there were three congruent segments and three incongruent segments. Therefore, six values (i = 1,…, 6) of any of the features were obtained from the signals recorded during these segments. Equation (2) was used to normalize each individual segment value dividing it by the sum of all six segment values.

$$Y_{si}' = \frac{Y_{si}}{\sum_{j=1}^{6} Y_{sj}} \tag{2}$$

In addition to these normalization steps meant to minimize the impact of individual subject responses on the affective state identification process, all features (GSRmean, BVPIBImean, STslope and PD) were scaled to the range of [0, 1] through max-min normalization, using Equation (3). This allowed the study of their performance with respect to a threshold that spanned a uniform range of possible values: [0, 1], through the corresponding Receiver Operating Characteristic (ROC) curves.

$$Y_{norm} = \frac{Y'_s - Y'_{s\min}}{Y'_{s\max} - Y'_{s\min}} \tag{3}$$

6. RECEIVER OPERATING CHARACTERISTIC ANALYSIS AND COMPARISON

Receiver Operating Characteristic (ROC) curves illustrate the trade-off made in a classifier between its "false positive rate" (which reflects the false alarm level, i.e., fraction of negative cases incorrectly classified as positive) and its "true positive rate" (i.e., the fraction of all positive cases correctly classified), as an adjustable threshold takes different values. The axes in the ROC are the false positive rate and the true positive rate, and each point of the ROC is defined for a specific setting of the threshold in the classifier. At the lowest sensitivity level (i.e., setting the threshold at the highest possible value of the detection signal) the classifier produces no false alarms but detects no positive cases. This is represented by the origin of the coordinate axes in the ROC diagram. As the sensitivity is increased, (i.e., as the threshold is lowered) the classifier detects more positive examples but may also start generating false alarms (false positives). Eventually the sensitivity may become so high (threshold set at the lowest possible value of the detection signal) that the classifier generates a detection (positive) for every input case. In this condition the classifier gets all positive cases right (true positive rate = 1), but it gets all the negative cases wrong, because it raises a false alarm on each negative case (false positive rate = 1). This corresponds to the top right-hand corner of the ROC. While all ROC curves "start" at the coordinate origin, (0,0) and "end" at the upper-right corner (1,1), the trajectory between these points followed by a given ROC, and consequently, the "Area under the ROC" are indicators of the discriminant power of the classification signal being thresholded. A "random classifier" (i.e., a process that produces uniformly distributed random numbers, without any relation to the input which is supposedly being classified) would display a ROC that follows approximately a 45^o diagonal ascent from (0,0) and (1,1). The "area under the ROC" (AUROC) would, therefore, be close to 0.5 (half the area of the 1.0-by-1.0 square). On the other hand, a classification system that produces a highly discriminating detection signal will have one or several threshold levels that map close to the upper-left corner of the ROC, at (0,1) indicating a high sensitivity (large true positive rate) and also a high specificity (low false positive rate). If that is the case, the AUROC will come close to encompassing the full 1.0-by-1.0 square. That is, the AUROC will approach the ideal value of 1.

The previous summary of ROC characteristics is useful in comparing the ROCs from the four selected classification signals, and should serve to emphasize the higher discriminant power found in the PD feature towards the differentiation of Stroop congruent and incongruent segments. The ROC curve for each of these detection functions (PD, GSRmean, BVPIBImean and STslope) has been estimated using the values for the 6(segments) x 32(subjects) = 192 segments analyzed in our experiments. Only half of these segments correspond

to "stressed" states induced by incongruent Stroop stimulation (ideal classifier output = 1), while the other half are known to be associated with "relaxed" (congruent Stroop) intervals (ideal classifier output = -1). Each point of the ROC curves is determined by comparing the value of the detection signal to a specific threshold level and determining which portion of the "positive classifier outputs" match the ideal (1) and which portion of the "negative classifier outputs" are in disagreement to the ideal output (-1). These "portions", expressed as fractional numbers, yield the coordinates of the ROC point (false positive rate, true positive rate) for the threshold value tested. The process was carried out using the ROC Matlab ® scripts provided by Dr. Gavin C. Cawley (University of East Anglia, Norwich, UK) in his web site http://theoval.sys.uea.ac.uk/matlab/default.html (accessed 01/2007).

Figures 5 through 7 show the ROC computed for the PD detection signal (solid line) alongside the ROC computed for each of the remaining detection signals (GSRmean, BVPIBImean and STslope, respectively), shown with dashed lines. It should be noted that the ROC curve for PD, exhibits a sharp slope from the coordinate origin, almost immediately reaching into high levels of true positive rate. Then the curve exhibits a number of intermediate points (threshold levels) for which the true positive rate is better than 0.8 while simultaneously having a false positive rate of less than 0.2. Accordingly, the area under this curve is large: AUROC_PD = 0.9647. Both the ROC curves for GSRmean and for BVPIBImean show a convexity that makes them depart from the random classification diagonal to some extent. However, their areas under the ROC are only moderately better than 0.5: AUROC_GSRmean = 0.6519 and AUROC_BVPIBImean = 0.6455.

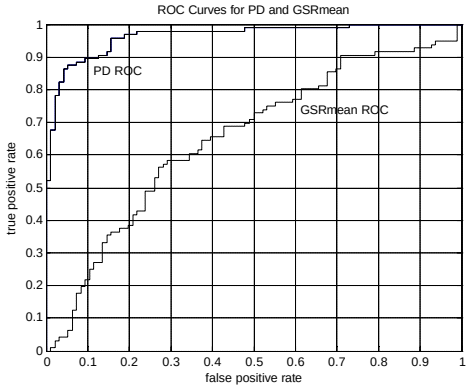

Figure 5: ROC curves for PD (AUROC = 0.9647) and GSRmean (AUROC = 0.6519)

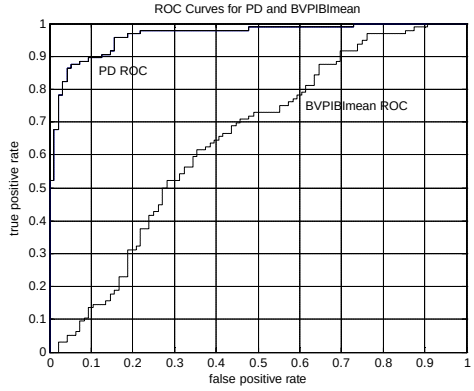

Figure 6: ROC curves for PD (AUROC = 0.9647) and BVPIBImean (AUROC = 0.6455)

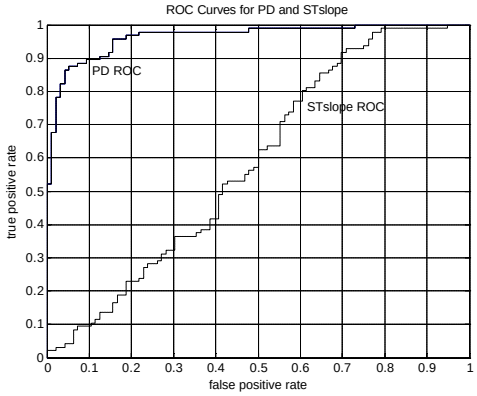

Figure 7: ROC curves for PD (AUROC = 0.9647) and STslope (AUROC = 0.5849)

The ROC curve for STslope shown in Figure 7 is actually very close to the random classification diagonal and, in general terms, follows a straight line at an angle just slightly larger than 45o. As such, the area under this curve is not much higher than ½ : AUROC_STslope = 0.5849 .

Figure 8 is offered as a visualization of the superior performance of the PD detection signal in terms of the area under the ROC curve (AUROC). PD clearly seems to provide a more powerful means for the discrimination of the congruent Stroop segments from their incongruent counterparts.

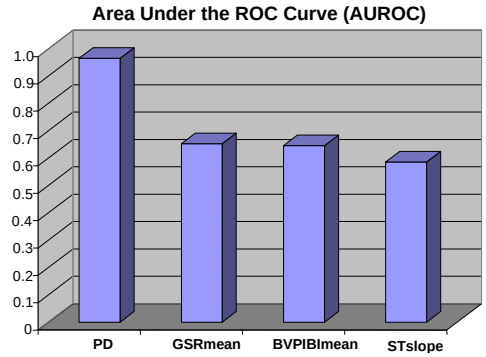

Figure 8: Areas under the ROC curves found for the four detection signals analyzed

7. CONCLUSIONS

This report has presented the comparison of four detection signals, derived from four corresponding physiological signals, for the identification of affective responses to congruent and incongruent Stroop stimulation. The four detection signals are GSRmean, BVPIBImean, STslope and PD. The comparative evaluation of these detection signals has highlighted the apparent enhanced potential of the average pupil diameter (PD), over the other three variables, to act as individual classification signals for the differentiation between stress and relaxation in computer users. It was observed that two signals derived from physiological variables traditionally used to detect autonomic changes (GSRmean and BVPIBImean) have only moderate discriminating power, which is inferior (in terms of the area under the ROC curve) to that of the PD signal. The STslope signal exhibited an even more limited potential for this detection problem, as the area under the corresponding ROC curve was the lowest of the four. While it must be acknowledged that the controlled conditions in which the study was performed are not necessarily representative of many real human-computer interactions, the potential displayed by the PD measurements warrant further investigation into mechanisms that may allow their use under more realistic scenarios.

ACKNOWLEDGMENTS

This work was sponsored by NSF grants CNS-0520811, IIS-0308155, HRD-0317692 and CNS-0426125. Ms. Ying Gao is the recipient of a Presidential Enhancement Assistantship from Florida International University.

REFERENCES

Barreto, A., and Zhai, J., "Physiologic Instrumentation for Real-time Monitoring of Affective State of Computer Users", WSEAS Transactions on Circuits and Systems, Vol. 3, Issue 3, pp. 496-501, 2004.

Chaffar S., and Frasson C., "The emotional Conditions of Learning". Proc. Int. FLAIRS Conference, AAAI Press, Clearwater , FL , USA , May 2005, pp. 201-206

Chaffar S., and Frasson C., "Predicting Learner's Emotional Response in Intelligent Distance Learning Systems". Proc. 19th Int. FLAIRS Conference, AAAI Press, Melbourne, FL , USA , May 2006. pp. 383-388.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G. 2001. "Emotion recognition in human-computer interaction". Signal Processing Magazine, IEEE 18:32-80

Martini, F.H., Ober W.C., Garrison C.W., Welch, K., and Hutchings, R. T., Fundamentals of Anatomy & Physiology, 5th Edition, Prentice-Hall, Upper Saddle River, NJ, 2001.

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

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

Siegle G.J., Steinhauer S.R, and Thase M.E., "Pupillary Assessment and computational modeling of the Stroop task in depression", Int. Journal of Psychophysiology 52 (2004) 63-76.

Steinhauer S.R., Siegle G.J., Condray R., and Pless M., "Sympathetic and parasympathetic innervation of papillary dilation during sustained processing", Int. Journal of Psychophysiology 52 (2004) 77-86.

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

Zhai, J., and Barreto, A., (a) "Stress Recognition Using Non-invasive Technology", Proc. 19th Int. FLAIRS Conference, AAAI Press, Melbourne, FL , USA , May 2006, pp. 395–400.

Zhai, J. and Barreto, A., (b) "Stress Detection in Computer Users Through Non-invasive Monitoring of Physiological Signals", Biomedical Sciences Instrumentation, Vol. 42, pp. 495-500, 2006.

Authorization and Disclaimer

Authors authorize LACCEI to publish the paper in the conference proceedings. Neither LACCEI nor the editors are responsible either for the content or for the implications of what is expressed in the paper.