Personalized Stress Detection from Physiological Measurements

Yuan Shi*, Minh Hoai Nguyen*, Patrick Blitz*, Brian French*, Scott Fisk*, Fernando De la Torre*,
Asim Smailagic*, Daniel P. Siewiorek*, Mustafa al' Absi†, Emre Ertin‡, Thomas Kamarck§, and Santosh Kumar¶
**Carnegie Mellon University* †*University of Minnesota-Duluth* ‡*The Ohio State University*
§*University of Pittsburgh* ¶*University of Memphis*
yuanshi@cmu.edu, minhhoai@cmu.edu, pblitz@andrew.cmu.edu, bfrench@andrew.cmu.edu
sfisk@andrew.cmu.edu, ftorre@cs.cmu.edu, asim@cs.cmu.edu, dps@cs.cmu.edu
malabsi@umn.edu, ertin.1@osu.edu, tkam@pitt.edu, skumar4@memphis.edu

***Abstract*—This paper describes a study on continuous, non-intrusive stress detection from physiological measurements, involving data collection, feature extraction, and model construction. We built a personalized stress detection model based on Support Vector Machines, and evaluated it on the collected data. Experimental results show that our model can detect stress with high precision.**

I. Introduction

Assessing stress in natural environment is beneficial for understanding the emotional behavior of human beings. Existing studies have shown that psychosocial stress can be assessed via self-report measurement, and such measurement is associated with biological and behavioral indices relevant to health. Examples include ambulatory blood pressure [1], carotid artery atherosclerosis [2], and cigarette smoking lapses [3]. Self-report measurement, however, requires a significant response cost on the participant, making frequent collection of such measurement impractical. If the stress assessment can be conducted on a continuous basis without requiring self-reports, it will significantly expand the usage of the assessment and benefit the research on stress and health.

Previous work has demonstrated the feasibility of detecting stress from physiological measurements. Such measurements can be acquired with minimal discomfort for the subject, and are useful in reflecting emotions [4]. One of the most relevant studies in this field was conducted by Healey and Picard [5], in which they presented methods for collecting and analyzing physiological data during real world driving tasks. They continuously recorded electrocardiogram, electromyogram, skin conductance, and respiration signals of drivers on a fixed route through Downtown Boston. Using Linear Discriminant Analysis [6], they showed that physiological measurements can predict mental stress with high accuracy. Healey and Picard's study, however, did not address physical stress [7] which is another important stress category.

In this work, we aim at a more complete study of both mental and physical stress. We collected the data through a controlled lab study containing different stressors and rest periods. For annotation, we collected Ecological Momentary Assessment (EMA) [8] results from the participants before and after each rest/stressor period. This way of acquiring stress annotation is different from Healey and Picard's approach which was based on judgement of human coders on video tapes of the driving. Their approach might be biased toward visually discernible effects, and may not well reflect the true stress state of the drivers.

For automatic stress detection, we trained personalized models using Support Vector Machines (SVMs) [9]. Experiments on the recorded data show that our model can achieve good precision at high detection rate. By combining advanced machine learning techniques and multi-sensory platforms to detect person's stress, we will proceed toward systems that can provide advice that is trusted. This research would enrich the quality of life technology capabilities.

II. Data Collection

The data was collected through a lab study which involved 22 subjects. Each subject in the study was exposed to a protocol containing four stressors and six rest periods. The four stressors were: one public speaking stressor, two mental arithmetic stressors, and one cold pressor stressor. These stressors represent the social, mental, or physical challenges that might lead to either mental or physical stress. Table I lists the sessions in our lab study.

For the public speaking task, participants were asked to prepare (four minutes) and deliver (four minutes) a speech while being videotaped. Participants were informed that they would be evaluated on poise, articulation, and style by two staff members present during the speech.

During the mental arithmetic tasks the subjects were asked to continuously add the digits of a three digit number and then add this sum to the original number. When a mistake was made, the participant was asked to go back to the previous correct number.

For the cold pressor test, the participant placed their hand in ice water while providing discomfort ratings to the test giver. This test lasted two minutes.

During the lab study, each subject wore an AutoSense[1]

[1]http://sites.google.com/site/autosenseproject/

Table I
SESSIONS IN THE LAB STUDY

Session	Approximate duration—minutes
Initial rest	30
Public speaking stressor	4 (preparation), 4 (talking)
Rest	5
Mental arithmetic I	4
Rest	5
Mental arithmetic II	4
Rest	5
Cold pressor	2
Final rest	20

platform on a chest band. The chest band contained the following sensors:

- Electrocardiogram (ECG): a two-lead ECG with electrodes was placed on the subjects' chest. The ECG signal was down-sampled to 60Hz.
- Galvanic Skin Resistance (GSR): the skin resistance was measured on the chest band at 10Hz.
- Respiration (RIP): the band carrying the sensor platform had a built-in resistor that was used to measure the chest expansion. The chest expansion, i.e., the electric impedance, was sampled at 60Hz.
- Temperature: the skin temperature was measured at 10Hz.

For stress annotation, we collected EMA interviews from the participants before and after each stressor/rest period. The questions and possible answers for each EMA interview are listed in Table II. To convert the EMA results to stress labels, we mapped YES, yes, no, NO to 3, 2, 1, 0, respectively, and the average of the answers to the third and fourth questions was calculated. If this value was larger than 1, the session immediately before the EMA interview was labeled as positive (stressed); otherwise, this session was labeled as negative (non-stressed).

Table II
EMA QUESTIONS AND POSSIBLE ANSWERS

Question	Possible answers
At the time of the prompt, how are you feeling:	
Cheerful?	YES yes no NO
Happy?	YES yes no NO
Frustrated/Angry?	YES yes no NO
Nervous/Stressed?	YES yes no NO
Sad?	YES yes no NO

III. FEATURE EXTRACTION

After the sensor data was collected, we extracted 26 types of features which were suggested to be relevant in related literature [5], [10], [11]. The features can be categorized in the following groups:

- Heart Rate features: HR, DHR, and DHR^2;
- ECG features: RSA, MF, LF, 01Hz, 12Hz, 23Hz, 34Hz, LFMFHF, and HFLF;
- Respiration features: RP, DRP, and RA;
- Skin Conductance features: SCL, DSCL, SCL^2, and SCLMAD;
- Galvanic Skin Response features: SRR, SRA, GSRD, and GSRA;
- Temperature features: T, DT, and DT^2.

The full list of all features are given in Table III. All features were calculated as the mean, standard deviation, or integration of sensor responses over a 60-second window. A feature vector was calculated every 20 seconds to form a data sample, which was labeled either positive or negative depending on the corresponding EMA responses. The number of positive and negative samples for 22 subjects are summarized in Table IV.

Table III
FEATURES EXTRACTED FROM SENSOR DATA

Feature	Description	Sensor
HR	mean heart rate	ECG
DHR	heart rate deviation	ECG
DHR^2	heart rate deviation squared	ECG
RSA	respiratory sinus arrhythmia, integration over the HF Band (0.15-0.5 Hz)	ECG
MF	integration over the power of the MF Band (0.09-0.15Hz)	ECG
LF	integration over the power of the LF Band (0.00-0.09HZ)	ECG
01Hz	sum of energy in 0-0.1 Hz Band	ECG
12Hz	sum of energy in 0.1-0.2 Hz Band	ECG
23Hz	sum of energy in 0.2-0.3 Hz Band	ECG
34Hz	sum of energy in 0.3-0.4 Hz Band	ECG
LFMFHF	(LF + MF)/HF	ECG
HFLF	ratio of sum of LF / HF	ECG
RP	mean respiration period (time between two respiration cycles)	RIP
DRP	deviation of the respiration period	RIP
RA	mean respiration amplitude	RIP
SCL	mean skin conductance level	GSR
DSCL	skin conductance level deviation	GSR
DSCL^2	skin conductance level deviation squared	GSR
SCLMAD	mean absolute deviation of the skin conductance level	GSR
SRR	number of GSR responses	GSR
SRA	amplitude of GSR responses in a window	GSR
GSRA	sum of the area of GSR responses in a window	GSR
GRSD	sum of the duration of GSR responses in a window	GSR
T	mean temperature	Temperature
DT	temperature deviation	Temperature
DT^2	temperature deviation squared	Temperature

IV. STRESS DETECTION MODEL

This section describes our stress detection model which extends SVMs to incorporate person-specific and temporal information.

A. Support Vector Machines

Given a set of training samples $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n \in \mathbb{R}^d$, and their corresponding labels $y_1, y_2, \cdots, y_n \in \{-1, +1\}$,

Table IV
NUMBERS OF POSITIVE AND NEGATIVE SAMPLES FROM 22 SUBJECTS

Subject	#Positive / #Negative	Subject	#Positive / #Negative
1	8 / 131	12	25 / 148
2	5 / 188	13	0 / 201
3	0 / 205	14	14 / 76
4	23 / 183	15	23 / 191
5	45 / 133	16	26 / 173
6	19 / 171	17	1 / 156
7	36 / 154	18	28 / 159
8	143 / 60	19	29 / 143
9	21 / 165	20	39 / 168
10	0 / 89	21	36 / 96
11	47 / 163	22	0 / 137

SVMs first map the input data to a feature space via the feature mapping $\varphi(\cdot)$, and seek a separating hyperplane with the maximum margin [12], [9]:

$$\begin{aligned} \underset{\mathbf{w},b}{\text{minimize}} \ & \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{n}\xi_i, \\ \text{s.t.} \ & y_i(\mathbf{w}^T\varphi(\mathbf{x}_i)+b) \geq 1-\xi_i \quad \forall i, \\ & \xi_i \geq 0 \quad \forall i. \end{aligned} \tag{1}$$

where $\mathbf{w}$ and b are the parameters of SVMs, $\{\xi_i\}$ are slack variables for penalizing the constraint violation, and C is a parameter balancing the trade-off between a large margin and less constraint violation.

B. Incorporating Person-specific Information

Although SVMs are state-of-the-art machine learning techniques and have been shown to yield excellent performance in many applications, their generalization capability might be limited if the inter-subject variation is large. As shown by Healey in [13], physiological responses can vary significantly between subjects. We therefore propose to investigate an approach that incorporates person-specific information, and builds personalized stress detection model to overcome this challenge.

Let r be the number of subjects, $\{(\mathbf{x}_i^p, y_i^p)\}_{i=1}^{n_p}$ the set of samples from subject p. Our personalized SVM formulation is:

$$\begin{aligned} \underset{\mathbf{w},b}{\text{minimize}} \ & \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{p=1}^{r}\sum_{i=1}^{n_p}\xi_i^p, \\ \text{s.t.} \ & y_i^p(\mathbf{w}^T\varphi(\mathbf{x}_i^p,\boldsymbol{\theta}^p)+b) \geq 1-\xi_i^p \quad \forall p\,\forall i, \\ & \xi_i^p \geq 0 \quad \forall p\,\forall i. \end{aligned} \tag{2}$$

Here $\boldsymbol{\theta}^p$ is a personalized parameter for subject p.

Previous studies [14], [15], [5] have shown that the changes in physiological measurements are more indicative for the transition of mental states than the absolute measurement values. Based on these studies, we propose a person-specific feature mapping to capture the deviation of physiological measurements:

$$\varphi(\mathbf{x}_i^p, \boldsymbol{\theta}^p) = \varphi(\mathbf{x}_i^p - \boldsymbol{\theta}^p). \tag{3}$$

Here $\boldsymbol{\theta}^p$ is person-specific parameter, which is estimated as the mean of all negative samples from subject p. To some extent, this is a *neutral* state; its role is analogous to the use of the neutral face in facial expression analysis [16].

C. Incorporating Temporal Information

The features described in Sec. III are calculated based on a 60-second window, which might not be long enough to capture the dynamics of stress events. Therefore, we propose to explore an approach that performs stress detection based on longer temporal extent by pooling information from consecutive windows. Let us refer to the features extracted in the 60-second window as frame-based. For longer temporal modeling, we consider segment-based features which are computed as the statistics (minimum, maximum, mean, median, and standard deviation) of frame-based features in the segment, as shown in Fig. 1.

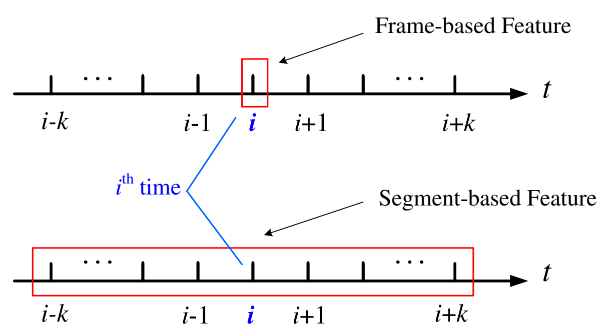

Figure 1. Frame-based and segment-based features. The segment-based features are computed by pooling statistics of frame-based features in the enclosing segment.

V. EVALUATION

To evaluate the detection performance, we used precision-recall values, instead of the more common ROC metric because the latter is designed for balanced binary classification rather than detection tasks [17]. Let tp, tn, fp, fn denote the numbers of true positives (i.e., samples correctly predicted as stress), true negatives, false positives, and false negatives, respectively. The precision and recall are defined as:

$$precision = \frac{tp}{tp+fp}, \quad recall = \frac{tp}{tp+fn} \tag{4}$$

Due to the limited amount of data, we evaluated the stress detector using leave-one-subject-out cross validation. Table V shows the the results of the generic model (without personalization) and the personalized one, both using frame-based features. Here we show the average precisions and standard errors at 80% recall. We experimented with two types of kernels: Linear and Radial Basis Function (RBF) kernels. As can be seen, incorporating person-specific information significantly improved the performance, 7% and 11% for Linear and RBF kernels, respectively.

Table V
AVERAGE AND STANDARD ERRORS OF PRECISION VALUES OF THE GENERIC AND PERSONALIZED MODELS AT 80% RECALL. BEST RESULTS ARE PRINTED IN BOLD.

kernel type	generic	personalized
Linear	0.56±0.054	**0.60±0.065**
RBF	0.56±0.053	**0.62±0.064**

Table VI demonstrates the effect of incorporating temporal information. We trained two personalized models, one with frame-based features and the other with segment-based features. We used RBF kernels for both models, and set the segment length to three minutes. As can be observed, the segment-based model significantly outperformed the frame-based one.

Table VI
AVERAGE AND STANDARD ERRORS OF PRECISION VALUES OF PERSONALIZED MODELS USING FRAME-BASED AND SEGMENT-BASED FEATURES AT 80% RECALL.

frame-based	segment-based
0.62±0.064	**0.68±0.073**

Figure 2 shows the precision–recall curve of the stress detection model with personalization, segment-based features, and RBF kernel. Even at 100% recall, the model can still achieve high precision over 50%.

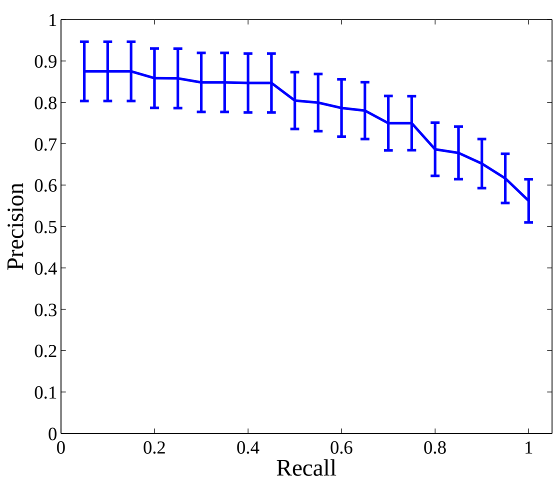

Figure 2. Precision–recall curve of our stress detection model

The tunable parameters of our models are C and γ of the kernels used in SVMs. In our experiments, these parameters were tuned using cross-validation on the training data.

VI. CONCLUSIONS

In this paper, we have presented SVM-based models for detecting stress from physiological measurements. Experimental results show that our models can detect stress at high precision and recall values, especially when personalized information is used. To our knowledge, this is the first stress detection model that incorporates person-specific information. Future work includes: i) validating the model through a field study; and ii) enabling online adaptation based on self-report measurement and contextual information.

ACKNOWLEDGEMENTS

This material is based upon work supported by the NIH grants on Real-Time Interview Technologies No.U01 023821 and Autosense project No. U01 DA023812, related NIH Opportunity fund grant on Personalized Stress Inference in Autosense, National Science Foundation under Grant No. EEEC-540865, Pennsylvania Infrastructure Technology Alliance (PITA), collaboration among Commonwealth of Pennsylvania, Carnegie Mellon and Lehigh University. We would like to thank Dr. Marcia Scott for her very valuable contributions to the project. Andrew Raij, Amin Ahsan Ali, Somnath Mitra, Motohiro Nakajima, Kurt Plarre, Mahbubur Rahman, and Siddharth Shah deserve our special thanks for their research and implementation efforts throughout this project.

REFERENCES

[1] T. Kamarck, D. Janicki, S. Shiffman, D. Polk, M. Muldoon, L. Liebenauer, and J. Schwartz, "Psychosocial demands and ambulatory blood pressure: A field assessment approach." *Physiology & Behavior*, vol. 77, no. 4-5, pp. 699–704, 2002.

[2] T. Kamarck, M. Muldoon, S. Shiffman, and K. Sutton-Tyrrell, "Experiences of demand and control during daily life are predictors of carotid atherosclerotic progression among healthy men." *Health Psychology*, vol. 26, no. 3, pp. 324–332, 2007.

[3] S. Shiffman and A. Waters, "Negative affect and smoking lapses: A prospective analysis," *Journal of Consulting and Clinical Psychology*, vol. 72, pp. 192–201, 2004.

[4] K. Kim, S. Bang, and S. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical and Biological Engineering and Computing*, vol. 42, no. 3, pp. 419–427, 2004.

[5] J. Healey and R. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, 2005.

[6] P. Lachenbruch and M. Goldstein, "Discriminant analysis," *Biometrics*, vol. 35, no. 1, pp. 69–85, 1979.

[7] N. Wallen, C. Held, N. Rehnqvist, and P. Hjemdahl, "Effects of mental and physical stress on platelet function in patients with stable angina pectoris and healthy controls," *European Heart Journal*, vol. 18, no. 5, p. 807, 1997.

[8] A. Stone, S. Shiffman, A. Atienza, and L. Nebeling, *The Science of Real-Time Data Capture: Arthur Stone*, 1st ed. New York: Oxford University Press, 2007.

[9] B. Schölkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. the MIT Press, 2002.

[10] S. Kreibig, F. Wilhelm, W. Roth, and J. Gross, “Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films,” *Psychophysiology*, vol. 44, no. 5, pp. 787–806, 2007.

[11] T. Oka, K. Oka, and T. Hori, “Mechanisms and mediators of psychological stress-induced rise in core temperature,” *Psychosomatic Medicine*, vol. 63, no. 3, p. 476, 2001.

[12] V. Vapnik and V. Vapnik, *Statistical Learning Theory*. Wiley New York, 1998.

[13] J. A. Healey, “Wearable and Automotive Systems for Affect Recognition from Physiology,” Ph.D. dissertation, 2000.

[14] J. Cacioppo, G. Berntson, J. Larsen, K. Poehlmann, T. Ito *et al.*, “The psychophysiology of emotion,” *Handbook of Emotions*, pp. 119–142, 1993.

[15] G. Wilson, J. Lambert, and C. Russell, “Performance enhancement with real-time physiologically controlled adaptive aiding,” in *Human Factors and Ergonomics Society Annual Meeting Proceedings*, vol. 44, no. 13. Human Factors and Ergonomics Society, 2000, pp. 61–64.

[16] P. Michel and R. El Kaliouby, “Real time facial expression recognition in video using support vector machines,” in *Proceedings of the 5th International Conference on Multimodal Interfaces*. ACM, 2003, p. 264.

[17] S. Agarwal, A. Awan, and D. Roth, “Learning to detect objects in images via a sparse, part-based representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1475–1490, 2004.