

Activity-Aware Mental Stress Detection Using Physiological Sensors

Feng-Tso Sun¹, Cynthia Kuo^{1,2}, Heng-Tze Cheng¹, Senaka Buthpitiya¹,
Patricia Collins¹, and Martin Griss¹

¹ Carnegie Mellon University
{lucas.sun,hengtze.cheng,senaka.buthpitiya,
patricia.collins,martin.griss}@sv.cmu.edu
² Nokia Research Center
cynthia.kuo@nokia.com

Abstract. Continuous stress monitoring may help users better understand their stress patterns and provide physicians with more reliable data for interventions. Previously, studies on mental stress detection were limited to a laboratory environment where participants generally rested in a sedentary position. However, it is impractical to exclude the effects of physical activity while developing a pervasive stress monitoring application for everyday use. The physiological responses caused by mental stress can be masked by variations due to physical activity.

We present an activity-aware mental stress detection scheme. Electrocardiogram (ECG), galvanic skin response (GSR), and accelerometer data were gathered from 20 participants across three activities: sitting, standing, and walking. For each activity, we gathered baseline physiological measurements and measurements while users were subjected to mental stressors. The activity information derived from the accelerometer enabled us to achieve 92.4% accuracy of mental stress classification for 10-fold cross validation and 80.9% accuracy for between-subjects classification.

Keywords: Mental stress, electrocardiogram, galvanic skin response, physical activity, heart rate variability, decision tress, Bayes net, support vector machine, stress classifier.

1 Introduction

Stress is a physiological response to the mental, emotional, or physical challenges that we encounter. Immediate threats provoke the body's "fight or flight" response, or acute stress response [5]. The body secretes hormones, such as adrenaline, into the bloodstream to intensify concentration. There are also many physical changes, such as increased heart rate and quickened reflexes. Under healthy conditions, the body returns to its normal state after dealing with acute stressors.

Unfortunately, many of the stressors in modern life are ongoing. Chronic stress can be detrimental to both physical and mental health. It is a risk

factor for hypertension and coronary artery disease [22,12]. Other physical disorders, including irritable bowel syndrome (IBS), gastroesophageal reflux disease (GERD), and back pain, may be caused or exacerbated by stress [16]. Chronic stress also plays a role in mental illnesses, such as generalized anxiety disorder and depression [11].

Chronic stress is difficult to manage because it cannot be measured in a consistent and timely way. One current method to characterize an individual's stress level is to conduct an interview or to administer a questionnaire during a visit with a physician or psychologist. This method provides only a momentary snapshot of the individual's stress level, as most individuals cannot accurately recall the history of the ebb and flow of their stress symptoms [3].

Continuous monitoring of an individual's stress levels is essential for understanding and managing personal stress. A number of physiological markers are widely used for stress assessment, including: galvanic skin response, several features of heart beat patterns, blood pressure, and respiration activity [31,?]. Fortunately, miniaturized wireless devices are available to monitor these physiological markers. By using these devices, individuals can closely track changes in their vital signs in order to maintain better health.

Measuring physiological signals during everyday activity is more difficult than in a rigorous laboratory environment. First, the physiological responses caused by mental stress can be masked by variations due to physical activity [1]. For example, people may have higher heart rate when standing than when sitting. Heart rate may also increase when people are mentally stressed. Hence, using heart rate alone as an indicator to detect mental stress may lead to misclassification. Second, signal artifacts caused by motion, electrode placement, or respiratory movement affect the accuracy of measured recordings. Third, it is also difficult to determine the ground truth of a user's stress level when labeling training data in mobile environment. These factors increase the difficulty of developing a pervasive mental stress detection application for everyday use.

We introduce an activity-aware, multi-modal system that combines accelerometer, ECG, and GSR information to differentiate between physical activity and mental stress. We conducted a user study with 20 participants across three different physical activities: sitting, standing, and walking. With activity information derived from the accelerometer, we achieved 92.4% accuracy for 10-fold cross validation and 80.9% accuracy for between-subject's classification.

In the next section, we describe how we can measure the body's responses to mental stress. Next, we discuss prior work on stress detection. Section 4 describes our experimental protocol and our physiological feature extraction and classification methods. Experimental results are presented in Section 7.

2 Background

The autonomic nervous system (ANS) regulates the body's major physiological activities, including the heart's electrical activity, gland secretion, blood pressure, and respiration. The ANS has two branches: the sympathetic nervous

system (SNS) and the parasympathetic nervous system (PNS). The SNS mobilizes the body's resources for action under stressful conditions. In contrast to the SNS, the PNS relaxes the body and stabilizes the body into steady state.

2.1 Heart Rate Variability (HRV) and Stress

Under acute stress, the SNS increases heart rate, respiration activity, sweat gland activity, etc. After the stress has passed, the PNS reverses the stress response [17]. Since the ANS controls the heart, measuring cardiac activity is an ideal, non-invasive means for evaluating the state of the ANS.

An ECG is a recorded tracing of the electrical activity generated by the heart. Figure 1 shows a P wave, a QRS complex, and a T wave in the ECG. The P wave represents atrial depolarization, the QRS represents ventricular depolarization, and the T wave reflects the rapid repolarization of the ventricles [8]. The R-R interval is the time interval between two R peaks and is used to calculate heart rate.

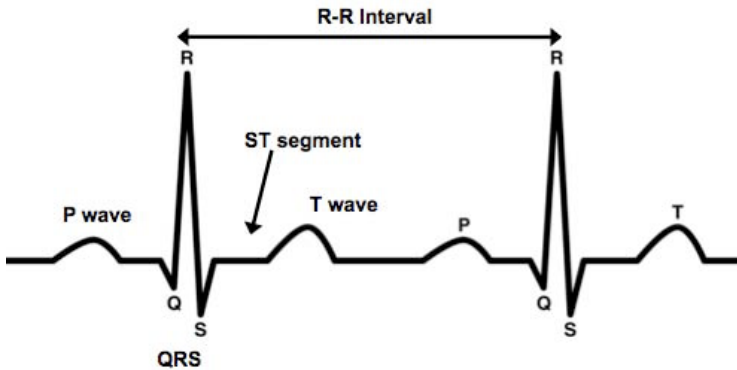


Fig. 1. Electrocardiogram sample

Heart rate variability (HRV) refers to the beat-to-beat variation in the R-R interval. HRV analysis can be categorized into time-domain and spectral-domain analysis. Several time-domain parameters include:

- mean HR: mean heart rate (beats per minute);
- mean RR: mean heartbeat interval (ms);
- SDNN: standard deviation of RR-intervals between normal beats;
- RMSSD: root mean square of the difference between successive RR-intervals; and
- pNN50: the percentage of heartbeat intervals with a difference in successive heartbeat intervals greater than 50 ms.

Three widely used components can be found in HRV power spectrum:

- LF (0.04-0.15 Hz): a low-frequency component that is mediated by both the SNS and PNS;
- HF (0.15-0.4Hz): a high-frequency component mediated by the PNS; and
- LF/HF: LF to HF ratio that is used as an index of autonomic balance.

2.2 Galvanic Skin Response (GSR) and Stress

GSR is a measure of the electrical resistance of the skin. A transient increase in skin conductance is proportional to sweat secretion[6]. When an individual is under mental stress, sweat gland activity is activated and increases skin conductance. Since the sweat glands are also controlled by the SNS, skin conductance acts as an indicator for sympathetic activation due to the stress reaction.

The hands and feet, where the density of sweat glands is highest, are usually used to measure GSR. There are two major components for GSR analysis. Skin conductance level (SCL) is a slowly changing part of the GSR signal, and it can be computed as the mean value of skin conductance over a window of data. A fast changing part of the GSR signal is called skin conductance response (SCR), which occurs in relation to a single stimulus. Widely used parameters for GSR include the amplitude and latency of SCR and average SCL value[2].

3 Related Work

The validity of using ECG and GSR measurements in mental stress monitoring has been demonstrated in both psychophysiology and bio-engineering. HRV analysis based on ECG measurement is commonly used as a quantitative marker describing the activity of the autonomic nervous system during stress. For example, Sloten et al. conclude that the mean RR is significantly lower (i.e., the heart rate is higher) with a mental task than in the control condition while pNN50 is significantly higher in the control condition than with a mental task [26].

Also, conventional short-term HRV features (e.g., a 5-minute sample window) may not capture the onset of acute mental stress for a mobile subject. Salahuddin et al. noted that HR and RR-intervals within 10 sec, RMSSD and pNN50 within 30 sec, high frequency band (HF: 0.15 to 0.4 Hz) within 40 sec, LF/HF, normalized low frequency band (LF: 0.04 to 0.15 Hz), and normalized HF within 50 sec can be reliably used for monitoring mental stress in mobile settings [23]. Hence, mental stress can be recognized with most HRV features calculated within one minute.

Boucsein provided an extensive coverage of early research of GSR related to stress [2]. He showed that slowly changing SCL and SCR aroused by specific stimulus are sensitive and valid indicators for the course of a stress reaction. Setz et al. demonstrated the discriminative power of GSR in distinguishing stress caused by a cognitive load and psychosocial stress by using a wearable GSR device in an office environment [25]. In this study, analysis of the data showed that the distributions of the SCL peak height and the SCR peak rate carry information about the stress level of a person.

Some research has used multiple physiological features to determine the existence of the subject's stress response. Zhai and Barreto applied an interactive "Paced Stroop Test," a psychological test of the subject's mental attention and flexibility, as a stimulus to elicit emotional stress in the subject [33]. The Paced Stroop Test requires the subject to select the font color of a word shown on the screen. The word itself names a potentially different color. The authors proposed

to extract features from the subject's physiological response (blood volume pulse, galvanic skin response, skin temperature and pupil diameter) during both the congruent phase (matching color name and font color) and incongruent phase (mismatching color name and font color). An example of the incongruent phase is shown in Figure 3. Three learning algorithms, Naive Bayes, Decision Tree, and Support Vector Machine (SVM), are used to classify relaxed and stressed states. The SVM classifier reached an accuracy of 90.1% with 20-fold cross validation.

Some experiments have been conducted in the real world. For instance, Healey and Picard measured drivers' stress reactions by monitoring multiple physiological signals, such as ECG, GSR, electromyogram (EMG), and respiration in a prescheduled route setting [10]. They used 5-minute intervals of data during the rest, highway, and city driving conditions to distinguish between three levels of driver stress. Heart rate and skin conductance provided the highest overall correlations with drivers' stress level across multiple drivers and driving days, reaching an accuracy of over 97%.

Most previous research considers distinguishing the physiological response to mental stress from subjects at rest. While developing mental stress monitoring algorithms in real-life ambulatory situations, it is crucial to take physical activity (e.g., walking) and posture (e.g., sitting or standing) into account. Cardiovascular variability is highly affected by changes in body posture and physical activity [30]. In Van Steenis et al.'s sample, subjects' mean HR increased significantly from a supine to sitting posture (from 66 to 77 bpm), from a sitting to standing posture (86 bpm), and from a standing posture to dynamic body movements (92 bpm). A major obstacle for ambulatory monitoring is that physiological dysregulation or emotion effects can be confounded by physical activity. Many physiological parameters, including heart rate, respiratory sinus arrhythmia, and skin conductance level, are strongly affected by both anxiety and exercise [32]. The daily routines involve different psychophysiological body activation characteristics. Kusserow et al.'s study showed that the physical-related and mental-related routines that are correlated with heart activity can be characterized and visualized as different activation patterns using RR-intervals and accelerometer data [13]. Schumm et al.'s work validated that it is possible to provoke and measure GSR with the startle event during different walking speeds [24]. However, the faster a person is walking, the more the peak distribution of GSR approaches a uniform distribution. Activity information is also helpful in ECG-based identity authentication area. The perturbation of the ECG signal due to physical activity is a major obstacle in applying the technology in real-world situations. Sriram et al.'s work presented a novel ECG- and accelerometer-based system that can authenticate individuals in an ongoing manner under various activity conditions [27].

Thus, physical activity distorts the result of mental stress detection in a mobile health monitoring scenario. In this paper, we compensate for the effects of physical activities by extracting a set of accelerometer features that characterize different physical activities along with ECG and GSR features. We hypothesize

that the accelerometer features provide the necessary auxiliary information for differentiate physical activity and mental.

4 Methodology

In this section, we describe the components of the wireless sensor system we used, the procedure of the experimental environment, and the segmentation of experimental dataset.

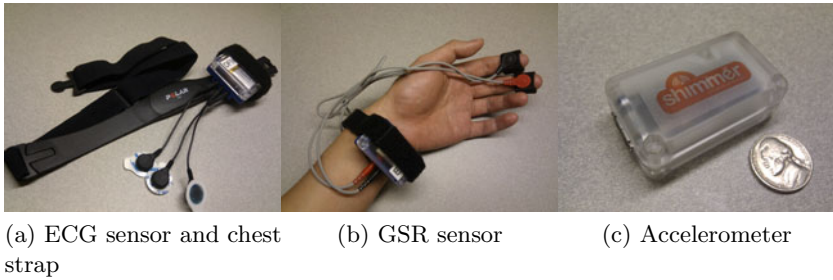


Fig. 2. SHIMMER sensors including ECG, GSR, and accelerometer

4.1 Wireless Sensor Network

We used the SHIMMER platform developed by Intel’s Digital Health Group. SHIMMER is a small wireless sensor platform with an integrated 3-axis accelerometer designed to support wearable applications. We also used SHIMMER’s ECG and GSR daughter boards for data acquisition. The sensor data from the ECG sensor and accelerometer were sampled at 100 Hz, and the data from the GSR sensor were sampled at 32 Hz. Data were transmitted to a PC via Bluetooth connectivity and saved to binary and comma-separated value files. We used three sensor nodes for the wireless sensor network configuration. Photos of the sensors are shown in Figure 2. The ECG sensor node was strapped to an elastic chest belt and three electrodes were placed on the body to form lead II and lead III¹ recording configurations. The GSR sensor was attached on a wrist band. Then, skin conductance was measured at the base of two fingers by measuring the electrical current that flowed as a result of applying a constant voltage. The third sensor node which was placed on the waist belt was used to collect accelerometer data.

¹ (Lead II is the voltage between the left leg (LL) electrode and the right arm (RA) electrode), and Lead III is the voltage between the (positive) left leg (LL) electrode and the left arm (LA) electrode.

4.2 Experimental Protocol

20 participants were monitored, 13 men and 7 women. Participants were students, faculty, and staff at our university. A computer application that randomly presented Stroop Color-word interference tests and mental arithmetic problems was provided as stressor. The Stroop test has been widely utilized as a psychological or cognitive stressor to introduce emotional response and autonomic reactivity [28]. Because participants would answer so many questions during the study, we added a variant of the Stroop test to prevent habituation, where participants were asked to select either color-name or font-color. The mental arithmetic is based on the Montreal Imaging Stress Task (MIST), consisting of two levels of difficulty under time pressure [7]. The mental arithmetic will adapt to participants' level or adjust the time limit in order to maintain an appropriate level of stress. The participants completed the mental tasks by interacting with a 19-inche touch-screen. Examples of two mental tasks are shown in Figure 3. When the participant provided an answer before the end of time limit, the feedback "correct" or "wrong" was displayed. The interface also shows the elapsed time and the participant's accuracy rate.

Participants were confronted with mental stress in each of three different conditions: sitting, standing, or walking. Each condition consisted of a baseline measurement with no stressor, measurement during the mental tasks, and a recovery segment:

1. Baseline segment (10 minutes): Listen to meditation music (in seated, standing, or walking position).
2. Mental task segment (10 minutes): Complete Stroop test and mental arithmetic under time pressure while seated, standing, or walking.
3. Recovery segment (10 minutes): Sit in a chair with closed eyes and listen to meditation music

All participants completed three conditions in random order. For the baseline and mental task segments, the participant had to complete the physical activity simultaneously. For example, when the participant was in the mental task segment of walking session, the participant was required to walk on the treadmill at 3 mi/hr and complete the mental task using the touch screen at the same time.

4.3 Data Collection

We collected sensor data from each participant for three physical activity conditions. The data were separated into six datasets. Hence, for each participant, we collected six data sets shown in Figure 4. The dataset for each segment contains 19200 GSR samples and 60000 ECG and accelerometer samples. The dataset for all 20 participants, including the six segments, is around 45 Megabytes.



(a) Stroop Test User Interface (b) Mental Arithmetics User Interface

Fig. 3. Screenshots of the stressor application

Baseline Segment (Sitting Only) 10 min	Stressed Segment (Mental Task + Sitting) 10 min	Recovery Segment (Sitting) 10 min
--	---	---

(a) Sitting condition

Baseline Segment (Standing Only) 10 min	Stressed Segment (Mental Task + Standing) 10 min	Recovery Segment (Sitting) 10 min
---	--	---

(b) Standing condition

Baseline Segment (Walking Only) 10 min	Stressed Segment (Mental Task + Walking) 10 min	Recovery Segment (Sitting) 10 min
--	---	---

(c) Walking condition

Fig. 4. Experimental conditions

5 Data Analysis

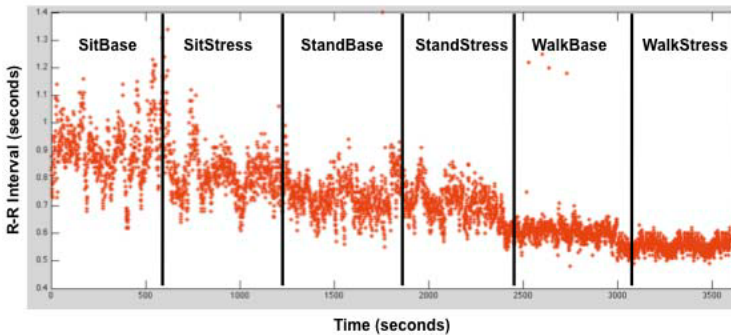
5.1 Feature Extraction

For each participant’s 60 minutes of data, we segment each channel of data into a 60-second window to obtain the data windows $\omega_1, \omega_2, \dots, \omega_{60}$. We denote F_i as the feature vector extracted from the data window ω_i and $F_i(j)$ is the j th feature in the feature vector F_i . We create a set of feature vectors F for each participant’s data set. Each segment in the experiment protocol has 10 feature vectors (e.g. $F_1 - F_{10}$ for SitBase, $F_{11} - F_{20}$ for SitStress, ..., $F_{51} - F_{60}$ for WalkStress).

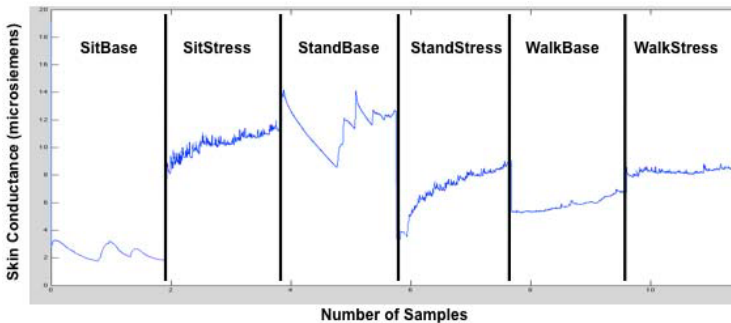
We chose a 60-second window for two reasons. First, the HRV features that we used in this study can distinguish between stressed and baseline segments using 60-second windows [23]. Second, the 60-second feature window reduces the impact of misclassified R-peaks by averaging HRV features within the window. All feature extraction algorithms are implemented in MATLAB. To eliminate

the artifacts caused by variations in electrode contact and physical motion, we applied both moving average and band-pass filtering techniques. For R-peak detection, we mainly adapted a derivative method [19] with modifications.

HRV analysis: HRV analysis methods can be categorized into time domain and spectral domain analysis. Time domain analysis is calculated directly from RR-intervals over the feature window. Examples of time domain features include mean value of the RR-interval (mean RR), standard deviation of the RR-interval (Std RR), mean value of the HR (mean HR), standard deviation of the HR (Std HR), RMSSD, and pNN50. Moreover, in the spectral domain methods, a power spectrum density (PSD) estimate is calculated for the RR interval series. Frequently used spectral measures are the very low frequency (VLF, 0.04 Hz), low frequency band (LF) and high frequency band (HF), and the ratio LF/HF. These spectral domain features are often interpreted as a measure of sympathovagal balance (autonomic state influence by the sympathetic and parasympathetic nervous system). We first calculated six time-domain features of HRV including mean RR, Std RR, mean HR, Std HR, RMSSD, and pNN50. Then, we applied a Fast Fourier Transform (FFT) to convert the time-domain RR-interval sequence



(a) RR interval data of a subject



(b) Skin conductance of a subject

Fig. 5. RR interval and GSR data in six experimental segments

to the power spectrum. The frequency components are used to calculate three spectral-domain features of HRV for each window: LF, HF, and LF/HF ratio.

GSR analysis: Due to the startle response (the physiological response of body to a sudden stimulus), the resistance of the skin can vary. The GSR can measure these subtle differences [29]. All GSR signals were filtered with a 256-point low-pass filter with 3Hz cutoff frequency to reduce noise. We calculated three GSR features: the total number of the startle responses in the segment, the sum of the response magnitude, and the sum of the response duration. These three features characterize the startle response, and Healey and Picard demonstrated their reliability [10]. Two additional features, mean and standard deviation of skin conductance level, are calculated over the feature window. Figure 5 shows the R-R interval and skin conductance recordings of a subject over six experimental segments.

Accelerometer analysis: Olguin and Pentland’s work indicated that an accelerometer placed on hip significantly helped classify activities such as sitting, running, crawling, and lying down [18]. Therefore, we placed one accelerometer on the waist belt close to the hip in order to maximize the difference of signal among sitting, standing, and walking activities. For each of the three axial dimensions, we calculated twelve features: mean value, standard deviation, energy, and correlation of each two axes. Table 1 lists the features derived from the ECG, GSR, and accelerometer data.

Table 1. Feature Vectors

	Sensor	Features
$f1-f9$	ECG	Mean RR, Std RR, Mean HR, Std HR RMSSD, pNN50, LF, HF, LF/HF ratio
$f10-f14$	GSR	Mean SCL, Std SCL, Total magnitude, Duration, and Number of startle responses
$f15-f26$	Accel	Mean of X, Y and Z axis Standard deviation of X, Y, and Z axis Energy of X, Y, and Z axis Correlation coefficient of XY, YZ, and ZX

5.2 Feature Normalization

Skin conductivity and heart rate signals are dependent on each individual’s initial physiological level. Even when the GSR or HR baseline level is measured from the same individual, these signals are likely day-dependent due to variations in physiology caused by diet or sleep, variations in mental state affected by mood, or variations in the sensor’s connectivity with skin [20]. Hence, to eliminate the intra-individual factor, we applied Equation 1 to each feature in feature vector set F . Since we conducted a short recording interval (one hour) for each participant, day-to-day variation caused by mood fluctuations is not considered in this study.

$$\begin{aligned}
 F(j)_{norm} &= \frac{F(j) - Z_{min}(j)}{Z_{max}(j) - Z_{min}(j)}, \text{ where} \\
 Z_{min}(j) &= \min\{F_i(j)\}, \forall i \in |F| \\
 Z_{max}(j) &= \max\{F_i(j)\}, \forall i \in |F|
 \end{aligned} \tag{1}$$

Equation 1 describes the normalization process for each feature. The first step is to subtract the minimum value from each feature such that the feature with minimum value becomes 0. Then, the feature values are divided by the overall range in six segments to make all the features lie between 0 to 1. The normalized feature values are fed to the classifiers described in the next section.

6 Stress Classification

We used the WEKA machine learning engine to train classifiers using various learning methods, including the J48 Decision Tree, Bayes Net, and support vector machine (SVM) for stress inference [9]. We divided the training data into two different sets in order to evaluate how activity information may influence the results of stress inference. One set of training data only includes the ECG- and GSR-related features while the second set also includes the accelerometer information. We also evaluated classification performance for between-subjects datasets and within-subject datasets.

6.1 Decision Tree

Decision Tree is a commonly used machine learning technique that uses a divide-and-conquer approach to classify testing data. During the learning stage, the tree structure is constructed. The tree structure has internal nodes and leaves. Internal nodes represent the test conditions while the leaves represent the classification results. We used a J48 Decision Tree for mental stress classification. Again, since we are interested in observing how the accelerometer information affects the accuracy of mental stress classification, we separated the training data into two sets. One dataset includes features extracted from accelerometer, ECG, and GSR recordings. The other dataset only consists of ECG and GSR features. When the Decision Tree is being constructed, the most informative feature (with a higher information gain) will be used near the root. Therefore, we are interested in observing the constructed decision tree structure to see if the accelerometer information is used in the test conditions and provides higher information gain.

6.2 Bayesian Network

We are also interested in using a Bayesian network structure to model the probabilistic relationships among physical and mental stress. Figure 6 shows two

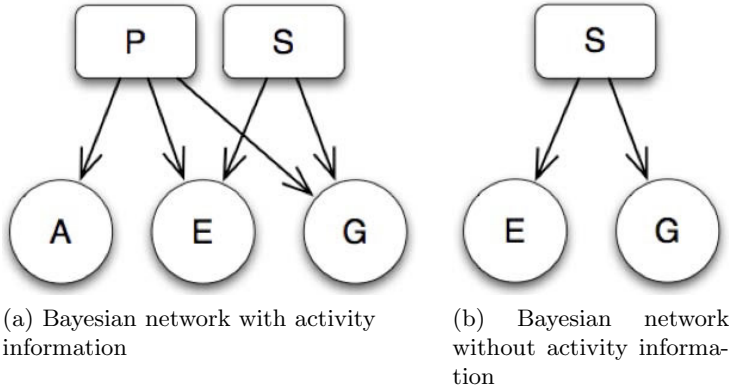


Fig. 6. Variable S represents a binary stress state and variable P represents the three physical activities. Variables A , E , and G are the accelerometer, ECG, and GSR features, respectively.

Bayesian network structures with and without considering the activity information to predict the existence of mental stress.

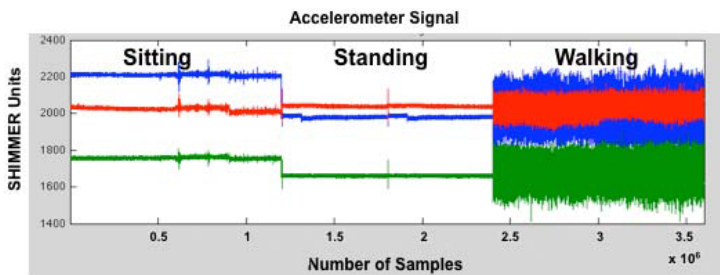
By comparing the inference results of the models in Figure 6a and in Figure 6b, we can investigate the effects of activity information in a mobile stress detection scenario.

In Figure 6a, $S \in \{\text{baseline, stressed}\}$ represents a binary stress state, and $P \in \{\text{sitting, standing, walking}\}$ represents the three physical activities. A , E , and G are the subsets of features defined in Table 1. $A = (a_1, \dots, a_{12})$ is a 12-feature vector corresponding to the accelerometer measure. $E = (e_1, \dots, e_9)$ is a 9-feature vector related to the HRV parameters. $G = (g_1, \dots, g_5)$ represents a set of 5 GSR-related features. Equation 2 shows the joint probability distribution encoded by the Bayesian network structure shown in Figure 6a. The Bayesian network structure shown in Figure 6b only uses physiological signals from ECG and GSR to infer the probability of the mental stress state.

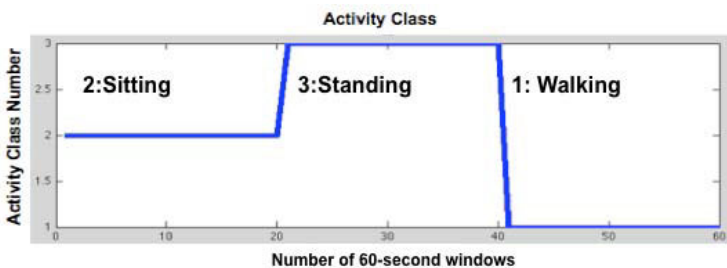
$$\begin{aligned}
 P(S, P, A, E, G) = \\
 P(S) \cdot P(P) \cdot P(A|P) \cdot P(E|P, S) \cdot P(G|P, S)
 \end{aligned}
 \tag{2}$$

We ran the K-Means clustering algorithm with the Euclidean distance metric on accelerometer data to automatically label variable P . Because we have three types of activities (sitting, standing, and walking), we set $K = 3$. Figure 7a shows an example of accelerometer raw data we collected from the experiment. Figure 7b shows the class of activity derived from Figure 7a using the K-Means algorithm.

The probability distribution of S conditioned on the three observed variables ($A = a$, $E = e$, and $G = g$) can be obtained by marginalizing the activity variable P as shown in Equation 3.



(a) Accelerometer data from one subject (red:x-axis, green:y-axis, and blue:z-axis)



(b) Activity classes derived from accelerometer data using the K-means algorithm

Fig. 7. Accelerometer data of a subject during three activities and the activity class derived by the K-Means algorithm

$$P(S|A = a, E = e, G = g) = \sum_{i=0}^{|p|} P(S, P = p_i | A = a, E = e, G = g) \quad (3)$$

The binary class of the stress state is estimated by maximizing the posterior probability in Equation 4.

$$s_{estimated} = \operatorname{argmax}_S \sum_{i=0}^{|p|} P(S, P = p_i | A = a, E = e, G = g) \quad (4)$$

An evaluation of cross validation for these two Bayesian networks is presented in Section 7.

6.3 Support Vector Machine

Support Vector Machine (SVM) is a classifier that performs classification by constructing a high-dimensional hyper-plane [4]. The constructed high-dimensional hyper-plane is optimized to separate the testing data into two classes. SVM also allows different types of kernel functions to transform testing data points into a higher dimensional space and make the transformed data easier to be classified. Since SVM has recently become a popular machine learning technique for classification, we are interested in investigating its performance with our testing dataset.

7 Experimental Results

In this section, we present the results from two experiments and a comparison of ECG and GSR feature efficacy. First, to investigate how a combination of features affect stress classification accuracy, we design four feature combinations from the measured accelerometer, ECG, and GSR data. Second, we test if the classifiers generalize across subjects by training our classifier on subset of subject data and testing our classifier on the remainder. Finally, we analyze the ECG and GSR features in six conditions across 20 participants.

7.1 Cross Validation with Different Feature Combinations

The first type of feature combination includes data measured from the accelerometer, ECG, and GSR. In each of the other three types of feature combinations, one feature is excluded. For each of the three classifiers described in Section 6, we evaluated its classification accuracy using these four types of feature combinations. Figure 8 plots the results of using 10-fold cross validation. For all of the three types of classifiers, excluding data recorded from the accelerometers degrades in the classification accuracy. The experimental results provide evidence to support our hypothesis that accelerometer data help the classification accuracy in a mobile stress detection scenario; physiological signals are both affected by physical activities and mental stress levels.

Furthermore, since heart rate is highly affected by the intensity of physical activities in our experiment, the classification results are even better for Bayesian network and SVM classifiers without including ECG features compared to the all-feature combination. Unlike ECG, the GSR features are good indicators to identify the presence of mental stress. When GSR features are excluded, the accuracy of each classifier decreases compared to the all-feature combination. We also found that the best classification accuracy (92.4%) is obtained from using the decision tree classifier with the all-feature combination. Moreover, the structure of the decision tree uses the energy of the x-axis from the accelerometer data as the root test condition. Several accelerometer features are also used as test conditions close to the root of the tree. It proves that activity information

provides higher information gain in the decision tree learning stage. Table 2 shows more detail of the cross validation results on 1200 samples. The grey cells highlight correctly recognized instances (true-positives).

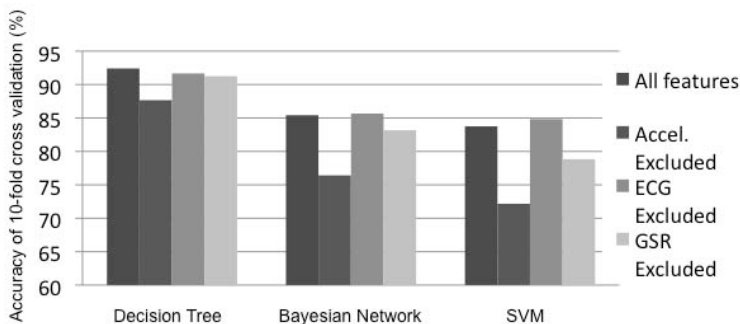


Fig. 8. Accuracy of the three classifiers using different feature combinations

Table 2. Confusion matrix for the combination of three classifiers and different feature combinations

		Classification schemes					
		Decision Tree		Bayesian Network		SVM	
Features		Baseline	Stressed	Baseline	Stressed	Baseline	Stressed
All features	Baseline	553	47	475	125	509	91
	Stressed	44	556	50	550	104	496
Accel. Excluded	Baseline	530	70	412	188	455	145
	Stressed	78	522	95	505	189	411
GSR Excluded	Baseline	539	61	431	169	461	139
	Stressed	44	556	33	567	115	485
ECG Excluded	Baseline	552	48	483	117	510	90
	Stressed	52	543	55	545	92	508

7.2 Between-subjects Experiment

We randomly selected a subset of our twenty subjects and used their data to train our classifier. Then, we tested the classifier on the remainder. For a given between-subjects classification setting, we repeated 10 times and calculated the average accuracy. To observe the effect on the size of the subset data, we changed the number of subjects in the training data set from 3, 6, ...to 18. The results shown in Figure 9 demonstrate that the SVM classifier outperforms the other two classifiers for between-subjects classification.

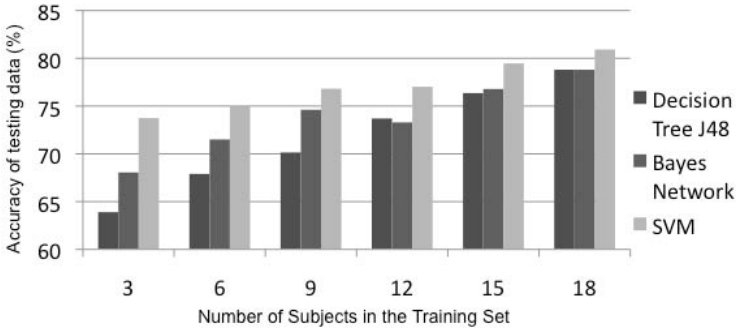


Fig. 9. Classification accuracy between subjects with the three classifiers

7.3 Comparison of HRV and GSR Parameters

Table 3 compares HRV features for the six experimental segments: (SitBase, SitStress, StandBase, StandStress, WalkBase, and WalkStress). Each HRV parameter is calculated by the average value across 20 participants for each segment. The mean RR interval decreases in all of the three mental task segments while the mean HR increases. The trend of mean RR and HR proves their efficacy in distinguishing mental stress across the three physical activities. The pNN50 feature also has a tendency to decrease from the baseline to the mental task segment. In the walking condition, pNN50 significantly decreases compared to the lower-intensity activities (sitting and standing). The standard deviation of RR is relatively high in walking condition and decreases the percentage of heartbeat intervals with difference in successive heartbeat intervals greater than 50 ms.

The last three rows in Table 3 are three spectral-domain parameters. The LF component is an indicator for both sympathetic modulations and cardiac vagal activity. It slightly increases from the baseline during all three activities. The HF component is considered to reflect parasympathetic modulations. There is a large increase in the HF component when the participants are walking. The imbalance of increase on LF and HF causes the LH ratio (LF / HF), a widely adapted index of sympathetic modulations, to decrease from sitting to walking. The HF increases during exercise has also been noted by other researchers [21].

From our analysis of all HRV parameters, we found that mean HR and RR are the most reliable features to recognize mental stress across three physical activities. The standard deviation of RR and HR did not demonstrate a coherent relation to the baseline and stressed segments. Spectral-domain parameters are sensitive to the physical activity conditions. Hence, this explains why excluding HRV features even increases in accuracy compared to the all-feature combination as shown in Figure 8.

Table 4 lists five GSR parameters for each segment. For each startle response, we can indicate its duration and magnitude. The total duration was calculated by accumulating the total elapsed time of the responses in the window. The total magnitude was measured by summing up the difference of the onset and the peak

Table 3. Comparison of HRV parameters in six conditions

HRV Parameters	Sit Base	Sit Stress	Stand Base	Stand Stress	Walk Base	Walk Stress
*Mean RR (ms)	887.59	814	752.07	722.43	586.03	562.94
Std RR (ms)	70.88	85.39	82.44	68.35	92.47	98.94
*Mean HR (bpm)	69.53	75.59	82.84	85.66	107.09	110.79
Std HR (bpm)	5.54	7.56	8.00	9.50	18.98	16.21
*pNN50 (%)	19.54	15.69	12.09	11.38	4.49	4.23
LF (%)	7.04	8.45	7.49	7.77	9.43	9.45
HF (%)	6.25	6.51	6.33	6.73	13.95	15.64
LH Ratio	1.34	1.51	1.45	1.48	0.67	0.71

of each startle response in window. The number of response occurrences over the one minute window was also recorded. Total duration, total magnitude, total occurrence of the responses, and mean GSR level illustrate an obvious increase from baseline to stressed segment. However, the standard deviation does not provide significant change between conditions.

8 Discussion and Conclusion

Previous mental stress studies were conducted in the laboratory with sedentary subjects. However, the controlled setting in a laboratory is not suitable for mobile mental stress monitoring because physical activity affects the measured physiological signals. The main goal of this study was to determine whether activity information can compensate for the interactive effects of mental stress and physical activity, which affect the accuracy of mental stress detection. Therefore, we conducted a user study in which participants completed baseline and mental task segments across three physical activities (sitting, standing, and walking).

This paper presented a multimodal approach to model the mental stress activation affected by physical activities using accelerometers, ECG, and GSR sensors. Our analysis showed that accelerometer data is necessary to improve mental stress detection in a mobile environment. We also noticed that the Decision Tree classifier has the best performance in our experiments using 10-fold cross validation. Decision Tree is recognized as one of the classification methods

Table 4. Comparison of GSR parameters in six conditions

GSR Parameters	Sit Base	Sit Stress	Stand Base	Stand Stress	Walk Base	Walk Stress
*Total duration(second)	3.17	14.30	4.16	13.15	13.72	16.32
*Total magnitude(μ Siemens)	0.79	2.04	0.75	3.32	1.69	1.97
*Total occurrence	1.09	6.58	3.13	6.37	5.63	7.47
*Mean GSR(μ Siemens)	4.69	4.83	6.19	6.97	6.42	7.22
Std GSR(μ Siemens)	0.62	0.53	0.62	0.71	0.63	0.52

with low computational complexity [14]. Therefore, the performance along with the low complexity of the Decision Tree classifier makes it a practical design choice for stress detection on mobile devices.

Furthermore, we also compared how physical activities and mental stress affect HRV and GSR parameters. We found that the GSR features are relatively independent of the three activities, even when participants were walking at a 3mi/hr. The between-subjects experiment demonstrated that we need to use up to 90% of subjects' data to achieve the classification accuracy of around 80%. It indicates that physiological signals tend to be user-dependent; hence, mental stress monitoring applications should also rely on personalized data in the training stage. We plan to further investigate the user-dependent attribute with more participants in the future. This study was limited to three specific activities and a relatively short recording time. The next step is to design a mobile platform enabling participants to wear sensors on a daily basis.

Our activity-aware scheme for mental stress detection can facilitate the development of many affective mobile applications using physiological signals (e.g., stress management, affective tutoring, and emotion-aware human computer interfaces). Including activity recognition techniques to interpret users' emotional states helps produce more feasible wearable sensors in everyday life.

References

1. Bernardi, L., et al.: Physical activity influences heart rate variability and very-low-frequency components in holster electrocardiograms. *Cardiovascular Research* 32(2), 234 (1996)
2. Boucsein, H.: *Electrodermal Activity*. Plenum, New York (1992)
3. Breslau, N., Kessler, R., Peterson, E.L.: Post-traumatic stress disorder assessment with a structured interview: reliability and concordance with a standardized clinical interview. *International Journal of Methods in Psychiatric Research* 7(3), 121–127 (1998)

4. Burges, C.J.: A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery* 2, 121–167 (1998)
5. Cannon, W.: *Bodily Change in Pain, Hunger, Fear and Rage: An Account of Recent Research into the Function of Emotional Excitement*. Appleton, New York (1915)
6. Darrow, C.: The rationale for treating the change in galvanic skin response as a change in conductance. *Psychophysiology* 1, 31–38 (1964)
7. Dedovic, K., Renwick, R., Mahani, N.K., Engert, V., Lupien, S.J., Pruessner, J.C.: The Montreal imaging stress task: Using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *Journal Psychiatry Neuroscience* 30, 319–325 (2005)
8. Dubin, D.: *Rapid Interpretation of EKG's*. Cover Publishing (2000)
9. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. *SIGKDD Explorations* 11 (2009)
10. Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems* 6(2), 156–166 (2005)
11. Herbert, J.: Fortnightly review: Stress, the brain, and mental illness. *British Medical Journal* 315, 530–535 (1997)
12. Holmes, S., Krantz, D.S., Rogers, H., Gottdiener, J., Contrad, R.J.: Mental stress and coronary artery disease: A multidisciplinary guide. *Progress in Cardiovascular Disease* 49, 106–122 (2006)
13. Kusserow, M., Amft, O., Troster, G.: Psychophysiological body activation characteristics in daily routines. In: *IEEE International Symposium on Wearable Computers*, pp. 155–156 (2009)
14. Lim, T.-S., Loh, W.-Y., Cohen, W.: A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning* 39 (2000)
15. Lundberg, U., et al.: Psychophysiological stress and EMG activity of the trapezius muscle. *International Journal of Behavioral Medicine* 1(4), 354–370 (1994)
16. Monnikes, H., Tebbe, J., Hildebrandt, M., et al.: Role of stress in functional gastrointestinal disorders. *Digestive Diseases* 19, 201–211 (2001)
17. T.F. of the European Society of Cardiology, The North American Society of Pacing, and Electrophysiology. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal* 17(2), 1043–1065 (1996)
18. Olguin, D., Pentland, Y.: Human activity recognition: Accuracy across common locations for wearable sensors. In: *Proc. 10th International Symposium Wearable Computer*, pp. 11–13 (2006)
19. Pan, J., Tompkins, W.J.: A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering* BME-32(3), 230–236 (1985)
20. Picard, R.W., Vyzas, E., Healey, J.: Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(10), 1175–1191 (2001)
21. Pichon, A., Bisschop, C., Roulaud, M., et al.: Spectral analysis of heart rate variability during exercise in trained subjects. *Medicine and Science in Sports and Exercise* 36, 1702–1708 (2004)
22. Pickering, T.: Mental stress as a causal factor in the development of hypertension and cardiovascular disease. *Current Hypertension Report* 3(3), 249–254 (2001)
23. Salahuddin, L., Cho, J., Jeong, M.G., Kim, D.: Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In: *29th Annual*

- International Conference of the IEEE, Engineering in Medicine and Biology Society, EMBS 2007, pp. 4656–4659 (August 2007)
24. Schumm, J., Bächlin, M., Setz, C., Arnrich, B., Roggen, D., Tröster, G.: Effect of movements on the electrodermal response after a startle event. In: Proceedings of 2nd International Conference on Pervasive Computing Technologies for Healthcare, Pervasive Health (2008)
 25. Setz, C., Arnrich, B., Schumm, J., La Marca, R., Troster, G., Ehlert, U.: Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on Information Technology in Biomedicine* 14(2), 410–417 (2010)
 26. Sloten, J.V., Verdonck, P., Nyssen, M., Hauelsen, J.: Influence of mental stress on heart rate and heart rate variability. In: International Federation for Medical and Biological Engineering Proceedings, pp. 1366–1369 (2008)
 27. Sriram, J.C., Shin, M., Choudhury, T., Kotz, D.: Activity-aware ECG-based patient authentication for remote health monitoring. In: ICMI-MLMI 2009: Proceedings of the 2009 International Conference on Multimodal Interfaces, pp. 297–304. ACM, New York (2009)
 28. Stroop, J.: Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18, 643–661 (1935)
 29. Tarvainen, M., Koistinen, A., Valkonen-Korhonen, M., Partanen, J., Karjalainen, P.: Analysis of galvanic skin responses with principal components and clustering techniques. *IEEE Transactions on Biomedical Engineering* 48(10), 1071–1079 (2001)
 30. Van Steenis, H., Tulen, J.: The effects of physical activities on cardiovascular variability in ambulatory situations. In: Proceedings of the 19th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society, vol. 1, pp. 105–108 (November 1997)
 31. Vrijkotte, T., et al.: Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension* 35(4), 880–886 (2000)
 32. Wilhelm, F.H., Pfaltz, M.C., Grossman, P., Roth, W.T.: Distinguishing emotional from physical activation in ambulatory psychophysiological monitoring. *Biomedical Sciences Instrumentation* 42, 458–463 (2006)
 33. Zhai, J., Barreto, A.: Stress detection in computer users based on digital signal processing of noninvasive physiological variables. In: 28th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society, EMBS 2006, pp. 1355–1358 (September 2006)