ReBreathe: A Calibration Protocol that Improves Stress/Relax Classification by Relabeling Deep Breathing Relaxation Exercises

Beena Ahmed, *Member, IEEE*, Hira Mujeeb Khan, *Member, IEEE*, Jongyong Choi, and Ricardo Gutierrez-Osuna, *Senior Member, IEEE*

Abstract—Training stress-prediction models is challenging due to the difficulty in reliably eliciting stress and relaxation responses in participants. For example, a task intended to elicit a relaxation response (e.g., deep breathing) can have the opposite effect depending on the participant's appraisal of and familiarity with the exercise. Including such instances in a training set undermines the accuracy of the resulting prediction model. This paper presents a technique, *ReBreathe*, to identify such instances based on respiratory patterns and determine their accurate *stress/relax* labels. We compared this relabeling approach against two labeling techniques: 1) nominal labels obtained from the experimental protocol and 2) labels obtained from subjective assessments. We then trained generalized estimating equation regression models to predict the resulting *stress/relax* labels from measures of heart rate variability and electrodermal activity. Training the model using protocol labels achieved a classification rate of 0.53 on participants not included in the training set. Relabeling the exercises based on each participant's subjective ratings increased classification rates but only marginally (0.61). In contrast, relabeling the exercises based on respiratory patterns increased classification rates to 0.88, or a four-fold reduction in error rates. These results illustrate the unreliability of protocol and subjective labels during *stress/relax* exercises and the potential benefits of *ReBreathe*.

Index Terms—Physiological signals, stress prediction, deep breathing, training protocols

1 Introduction

Today's fast-paced lifestyle exposes people to multiple and omnipresent stressors that can have adverse physiological effects on the human body [1], [2]. Effective stress management involves understanding how the body responds physiologically to stress and then learning how to regulate better our responses. The broad availability of wearable physiological sensors makes it feasible to monitor physiological correlates of stress over longer periods to predict stress levels; thus enabling improved diagnosis and early intervention [3]. Systems predicting stress from these physiological variables have been tested in the lab setting with some success [4], [5], [6], [7]. However the use of these systems is still limited, as it is hard to classify user-specific responses as either stressed or relaxed due to individual differences in collected physiological variables.

In particular, the accuracy of a stress prediction model depends greatly on being trained with correctly labeled, user-specific instances of stressful or relaxing events. When using protocol labels as ground truth, a premise is made that all participants experience stress in activities intended to induce stress and relaxation in activities intended to induce relaxation; however this premise may not always be valid due to individual differences. Standard activities that induce relaxation may, in some cases, elicit stress [8], depending on the participant's appraisal of and familiarity with the exercise. In an earlier work [9], we found significant variations in individual responses to standard stressful and relaxing activities. These variations illustrate the difficulty of the problem, both in terms of designing stress and relaxation elicitation protocols, as well as in taking these labels at face value. Self-reporting is typically used to qualify individual perceived stress, but it can be unreliable as it is a subjective measure [10]. The absence of a reliable ground truth to label stress states [7], necessitates an alternative methodology to identify accurate labels (i.e., *stress/relax*) for the activities in training protocols.

In this paper, we propose a calibration protocol, *ReBreathe*, which uses respiratory data to provide user-specific label correction for relaxation and stress. This correction can then be used to calibrate ambulatory stress prediction systems. Deep breathing, a proven relaxation technique [11], [12], [13], is included in *ReBreathe* to elicit relaxation in the participants. Our approach consists of using respiratory data to determine if participants adhere to a prescribed rate of six breaths per minute (hence relax) and relabeling the deep-breathing exercises as *stress/relax* accordingly. *ReBreathe* thus provides a relaxation reference [11], [12], [13] that takes into account individual variations

- *B. Ahmed and H.M. Khan are with the Electrical and Computer Engineering Program, Texas A&M University at Qatar, Doha, Qatar. E-mail: {beena.ahmed, hira.khan}@qatar.tamu.edu.*
- *J. Choi is with the Intel Coproration, Austin, TX. E-mail: goonyong@cse.tamu.edu.*
- *R. Gutierrez-Osuna is with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843–3112. E-mail: rgutier@cse.tamu.edu.*

Manuscript received 20 July 2014; revised 25 Apr. 2015; accepted 1 May 2015. Date of publication 21 July 2015; date of current version 6 June 2016. Recommended for acceptance by C. Lisetti.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TAFFC.2015.2459682

by determining whether participants perform deep breathing correctly to relax during these activities.

To validate *ReBreathe*, we conducted a series of experiments in which participants completed a calibration protocol consisting of various short-term stress-inducing activities alternated with repeated brief deep-breathing exercises. The goal of the experiments was to determine the physiological manifestation of short-term stress-inducing activities, not those of chronic stress. Throughout the experiments, we recorded the participants' physiological data via a wearable sensor system [3] and collected subjective ratings of stress levels via a questionnaire. We then compared the accuracy of our respiratory-based relabeling approach against the standard approaches used in the literature: 1) protocol-based labels and 2) labels obtained from self-reported subjective scores. Namely, we used the respiratory-based and the two standard labels each to train three corresponding generalized estimating equation (GEE) regression models using only heart rate variability (HRV) and electrodermal activity (EDA) features as independent variables. We implemented GEE regression models due to the repeated nature of the activities in *ReBreathe*. Respiratory features were utilized only for calibration in *ReBreathe* and not in the short-term stress prediction model. Our results indicate that relabeling deep breathing activities based on the participants' respiratory performance provided higher classification results with the GEE model when compared to results from models developed using the two standard labelling approaches.

The paper is organized as follows. Section 2 contains a brief overview of the need for the study and a summary of prior work in the area. Section 3 presents the three labeling approaches evaluated in this study as well as the experimental protocol and wireless sensor system used to collect physiological data. Section 4 discusses the collected physiological features and results of the GEE *stress/relax* prediction model. In the end, Section 5 contains a discussion of results and conclusions drawn.

2 Background

Stress arises from the reaction of the body to outside challenges, either physical or psychological. Lifestyle-related psychological stressors can be pervasive and persistent for long periods. This prolonged stress can be detrimental to health resulting in digestion and sleep disorders, obesity, cardiovascular diseases, immune system impairment, and psychological problems [14]. The World Health Organization declared stress the second most common health issue in the European Union, affecting one third of the employed population [15]. In Great Britain, 40 percent of the population has experienced health issues due to work-related stress [16], and more than 50 percent of the United States population suffers from the stress of balancing work and family life [17].

2.1 Assessment of Stress

The concept of stress, as first proposed by Selye in 1936, is now well known, though the author himself struggled unsuccessfully to find a satisfactory definition of stress due to its subjective nature [18]. As perceived stress can be both physical and psychological, assessing the stress level experienced by an individual is non-trivial. Though, it is possible to measure the physiological response to stress (e.g., heart rate, respiration, and cortisol), the impact of stress on these measures can vary considerably due to individual differences. Moreover, self-reporting scores [19], which provide a subjective assessment of the individual level of perceived stress, may also prove inaccurate due to differences in perceptions, participants' inability to recall the various activities, or their eagerness to adjust their responses to please the experimenter [20]. The association between physiological measures and subjective measures of chronic/acute stress has been well reported [21], [22], however its relationship with subjective measures of short-term stress is inconsistent and less significant [23], [24]. As the goal of wearable stress monitoring devices is to track the adverse effects of short-term stressors to prevent chronic/acute stress, the accurate quantification of short-term stress levels is vital.

2.2 Physiological Response to Stress

The response to stress is mediated by the activation of the sympathetic nervous system (SNS), one of the two branches of the autonomic nervous system (ANS) [25]. Another branch of the ANS, the parasympathetic nervous system (PNS), works in the opposite fashion, conserving energy. A number of physiological changes, including increased heart rate, pupil dilation, irregular breathing, increased electrodermal activity, and muscle tension, are recognized as reliable indices of SNS stimuli and stress [2]. All of these signals can be observed using wearable sensors [3]. Heart rate variability (i.e., variations in the beat-to-beat interval) indicates how well the ANS maintains equilibrium in the body. High values of HRV indicate that the SNS and PNS are working in proper balance as the person is in a relaxed state, whereas low HRV indicates that the SNS tone is dominant. As the SNS influence augments due to stress, the breathing rate also increases and becomes irregular [26]. Stress also affects respiratory sinus arrhythmia (RSA), a naturally occurring modulation of heart rate at the frequency of breathing that minimizes the work done by the heart [27]. Psychological stress influences the interaction of respiration and heart rate, disrupting the coherent oscillations of RSA, which in turn impacts HRV [28]. EDA reflects variations in electrical conduction of skin caused by perspiration occurring as a reaction to physiological and psychological arousal [29]. Respiratory parameters have been used in psychophysiology to index the effects of stress and emotion [30], [31] as psychological distress generally leads to increases in respiration rate [30] and minute volume, and a shift from abdominal to thoracic breathing [32].

2.3 Deep Breathing

A number of techniques have been applied to combat stress, including deep breathing, biofeedback, guided imagery, progressive muscle relaxation, cognitive behavioral therapy, and mindfulness-based stress reduction [33], [34]. Deep breathing is the cornerstone of all these techniques as it is effective in regulating ANS after acute stressful activities if performed correctly [12]; using the diaphragmatic muscles for breathing shifts the ANS towards the parasympathetic tone and thus induces relaxation [35]. Slow and deep breathing has also been found to lead to increases in

RSA and consequently HRV due to improvements in the synchronization between respiration and cardiac activity [28]. Deep breathing or diaphragmatic breathing is thus widely recommended by health professionals to help relieve stress [11], [36]. Successful deep breathing has been proven to relax participants in a range of different studies: healthy volunteers [12], [13], [37], pre-hypertensive women [38], medical students [39], [40], nursing students [41], hospital cleaners and bank employees [42], and athletes [43].

2.4 Physiological Stress Modelling

A number of devices have been developed to monitor physiological signals with validated relationships to stress [44], [45]. As the physiological response provides an incomplete picture of the stress state of an individual, behavioral patterns and self-reported psychological surveys have also been monitored [46]. The vast majority of experimental work done to validate these stress monitoring devices has been restricted to a controlled set of activities within lab settings. Stress labels are obtained based upon universal assumptions for all participants (e.g., mental arithmetic tasks are stressful and quiet time or listening to music relaxing for all participants [5], [47]. Models based on these labels include a Fisher's Least Square Linear classifier using EDA, HRV, EMG, and respiratory features [5], which reported accuracy rates of 79 percent with a five-fold cross-validation. On the other hand, a SVM classifier using blood volume pulse, EDA, skin temperature, and pupil diameter [6] and a fuzzy logic based system using only HR and EDA features [48] had higher accuracy rates of over 90 percent. However research is needed on developing models that account for variations in individual stress responses to these activities and how they impact the stress labels used. Further investigation is also needed to analyze variations in data collected over long periods of time and their correlation with stress [49].

In contrast with the extensive work done on physiological monitoring of stress, work on developing robust calibration protocols is limited. Some studies [4], [47] incorporate a calibration protocol consisting of a stress condition preceded by an initial rest period and followed by a recovery period. The rest period is used to obtain a baseline and assumes that participants are not stimulated in anticipation of the subsequent task. However, the assumption that all participants experience stress in a stress condition and relax in the rest period may not always be valid due to the variability in individual responses to the stimuli. Alternatively, other studies [46], [48], [49] use self-reported subjective scores as ground truth for calibration activities containing a range of stressors, which as discussed earlier can be inaccurate. Furthermore, participants have difficulty assessing stress levels of activities conducted in the lab, which, for ethical reasons, are moderate. As a result, using protocol labels or subjective ratings can be problematic for both of these methods as they make invalid assumptions about the predictability of the participants' responses.

3 Methods

3.1 Activity Labeling Approaches

A typical two-class supervised classification scenario consists of a training set $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$ containing N instances, where $x_i \in \mathbb{R}^d$ is a d-dimensional feature vector (e.g., physiological responses) and $y_i \in \mathcal{Y} = \{0, 1\}$ is a label (e.g., relaxed versus stressed). The task is to train a classification function $f \in \mathbb{R}^d \rightarrow \mathcal{Y}$ that generalizes to unseen data. However, due to the high level of individual variability, the true labels y_i are not known.

To identify the true labels y_i, our proposed calibration protocol, *ReBreathe*, involves using deep breathing (DB) exercises as a baseline activity in addition to short-term stress-inducing activities. We chose paced DB for two reasons: 1) its effectiveness in stress regulation (if performed correctly) is well-documented [12], [13], [37] and 2) it is possible to monitor breathing rate to determine if participants are following the protocol. If respiratory data were used directly without the DB protocol, it would be necessary to collect data over a longer period together with subjective data to identify relaxation periods.

We used *ReBreathe* to label the outcome of DB activities on a participant-by-participant basis. We then compared the obtained labels to those from the standard protocol-based and subjective score-based labeling approaches by employing the following three sets of labels to train three corresponding GEE regression short-term stress prediction models. Below are details of the labeling approaches:

3.1.1 Respiratory-Based Labeling (ReBreathe)

In this approach, we used respiratory data to assess whether participants had been able to perform their DB activities correctly. We then used this DB assessment to generate *stress/relax* labels of each activity. Namely, DB activities in which the power of the participants' respiratory signal corresponded to the prescribed rate of 0.1 Hz were relabeled as $y_i = 1$, (i.e., stressed) and the activities that did not were kept as $y_i = 0$ (i.e., relaxed). We used the power of the participants' respiratory signal within the 0.04-0.15 Hz band as a discriminating and label correction feature as participants breathing at rates significantly different from the prescribed 0.1 Hz would result in a lower power in the 0.04-0.15 Hz band. As *ReBreathe* was conducted in a lab setting, the data contained limited motion artifacts. The breathing rate was not controlled in the stressful activities in the protocol, hence the labels for the stress-inducing activities were kept at $y_i = 1$ (i.e., stressed). It is important the calibration protocol be conducted in a controlled environment, as in ambulatory settings the physiological data can become unreliable due to motion artifacts.

3.1.2 Protocol-Based Labeling

In this approach, we implemented standard protocol-based labelling [4], [5], [47]. We labeled stress levels according to the characteristics of the activity, thus assuming that all participants experienced stress while performing challenging mental activities and experienced relaxation during DB. As a result, all DB activities were labeled $y_i = 0$, and all stress-inducing activities labeled $y_i = 1$.

3.1.3 Subjective-Based Labeling

In this approach, we labeled activities according to the subjective scores provided by the participants [46], [49]. Namely, all (DB and stress) activities that received a

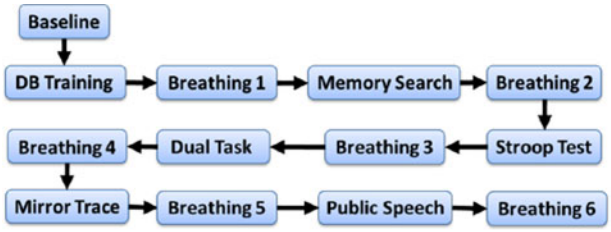

Fig. 1. Sequence of all the *stress/relax* activities performed by the participants during the experiment.

Fig. 2. Picture of participant wearing the custom wearable sensor system. The system consists of a holster unit, a wireless chest strap combining HRM and respiration sensors and wireless EDA module.

subjective score higher than or equal to a threshold were relabeled as $y_i = 1$, (i.e., stressed) and those that received scores below the threshold were kept as $y_i = 0$ (i.e., relaxed). The threshold used was the undecided (midpoint) subjective score of the seven-point Likert scale used by the participants (as detailed in Section 4.2).

3.2 Experimental Protocol

To validate *ReBreathe*, we conducted a series of experiments in which participants completed various short-term stress-inducing activities interleaved with repeated deep-breathing exercises, as illustrated in Fig. 1. Repeated DB activities were added to allow participants to recover from the stress-inducing activities. At the start of the experiment, participants were asked to sit quietly for two minutes to provide reference physiological signals. Next, participants were instructed on how to perform DB at a pace of six breaths per minute (0.1 Hz), i.e., to inhale for four seconds and then exhale for six seconds. They were then allowed to practice DB independently for five minutes. The participants performed DB five more times during the experiment, each time following a stress-inducing activity. The duration of these DB activities was three minutes.

After completion of each activity, participants rated how strongly they agreed or disagreed with the statement, 'The task I completed was extremely stressful' using a Likert scale as shown in Table 1; 1 corresponded to least stressful and 7 to most stressful.

Following our prior work [3], we used five different stress-inducing activities to reduce complacency due to repetition and ensure participants were challenged across a range of different skills:

- Memory search [50]: Participants memorized a set of words on the screen in a random sequence and identified them amongst a number of confounders in a limited time (5 min).
- Dual tracking [51]: Participants used a mouse to track a moving target in a square box on the computer screen and simultaneously left-click whenever one of three target letters appeared on the screen (5 min).
- Mirror tracing [52]: Participants manually traced a pattern on paper visible not directly but only by looking through a mirror placed at a strategic angle (5 min).
- Stroop Test [53]: Participants had to click on one of four buttons according to their ink color showing one of four words (red, green, blue, yellow) written in different ink colors to the button color (5 min), and
- Public speech [54]: Participants had to prepare a short speech on a topic provided (3 min), deliver the speech in front of a small audience (4 min), and address the audience's questions (3 min).

TABLE 1
Likert Scale Used to Rate Each Activity

1	Disagree completely
2	Disagree strongly
3	Disagree slightly
4	Un-decided
5	Agree slightly
6	Agree strongly
7	Agree completely

Twenty-five volunteers (ages 18-35 years) participated in the study. Prior approval for the study had been obtained from the TAMU Institutional Review Board. All of the volunteers (10 female, 15 male) were examined by a medical doctor to assess their suitability to participate in the study. Based on the medical results, 22 healthy people were selected for further experimentation. Three participants were excluded by the clinician due to pre-existing medical conditions that could either expose the participant to undue risk or affect the physiological data collected during the experiment (e.g., hypertension, diabetes). To identify if any participants were chronic stress sufferers, they were asked to complete the Perceived Stress Scale (PSS) before commencing the experiment [55]. All participants recorded scores less than 17 on the PSS, which is comparable to the mean PSS score in the US for college students less than 25 years (16.78) and thus not indicative of chronic stress (i.e. PSS $= 21+$) [56]. Studies were performed during the day between 8:00 am-4:00 pm. Participants were briefed on the experimental procedure to be followed and their written consent to participate was obtained prior to the experiment. The participants were not trained previously for any of the activities including the DB ones.

We collected physiological data using a custom wearable sensor system described in [3]; see Fig. 2. The system consisted of a single chest strap incorporating 1) a heart rate monitor strap (Polar® WearLink+®; Polar Electro Inc.) and a pressure-based respiration sensor (SA9311M; Thought Technology Ltd.); 2) small AgCl electrodes (E243; In Vivo Metric Systems Corp.) on the middle and index finger of the non-dominant hand to measure electrodermal activity EDA; and 3) a holster unit consisting of a data processing

TABLE 2
Brief Description of the Features Used for Analysis

Features	Description	Relationship to stress
Average of N-N beats **AVNN** [60], [61]	Time domain HRV—Mean of the time interval between normal sinus beats	↓ with stress
Root mean square of successive difference (msec) **RMSSD** [61]	Time domain HRV—exaggerated by irregular Heart rate	↓ with stress
%age difference between adjacent NN intervals greater than x msec **pNNx** [60], [61]	Time domain HRV—exaggerated by irregular heart rate	↓ with stress
HRV—High Frequency Component **HRV-HF** [61]	Frequency domain HRV—spectral power of NN intervals of 0.15-0.4 Hz	↓ with stress
Respiration—Low Frequency Component **RESP-LF** [62]	Spectral power of the signal between 0.04-0.15 Hz	↓ with stress
Skin Conductance Response **SCR** [63]	Mean of rapidly varying phasic response	↑ with stress

unit (Marvell^{TM} PXA270 400 MHz, 64 MB RAM; Gumstix, Inc.), a sensor hub (HRM receiver module—Polar RMCM01, Polar Electro Inc.; wireless transceiver—EZ430-RF2500, Texas Instruments Inc.), and a 3,000 mAh Lithium-polymer battery. A wireless sensor network was developed with the wireless nodes integrated into the chest band, EDA sensor and holster unit. Each wireless node consisted of a transceiver module, which created a star network topology for wireless transmission between the sensors (node) and the holster unit (hub).

Once the data was collected, we extracted several physiological indices from the heart rate, respiratory and electrodermal activity sensors as follows:

- Heart rate variability: We selected three time-domain measures (AVNN, RMSSD, PNNx) and one frequency-domain measure (HRV-HF); see Table 2. In the statistical family of pNNx, we selected x = 25 msec instead of pNN50 as it has been shown to provide more potent estimates of cardiac vagal tone [57]. The frequency domain analysis of HRV measures rhythmic oscillations of heart rate at different frequencies [58].
- Respiration: We computed the respiratory low frequency component (RESP-LF; 0.04–0.15 Hz) to determine if participants followed the prescribed respiratory rate of 0.1 Hz. We expected the participants' RESP-LF values to be high during the DB activities as their breathing rate would be slow, increasing dominance of the PNS.
- Electrodermal activity: Finally, we also extracted the skin conductance response (SCR) as a measure of EDA as it is a reliable indicator of stress [4] not influenced by the respiratory response. SCR responds rapidly to elicited stimuli, whereas skin conductance level (SCL) reflects slower changes unrelated to the stimuli indicative of a general level of arousal [59]. As in our experiment, the activities ranged from 3-5 minutes, we chose the phasic SCR instead of the tonic SCL. We used a regularized least-squares detrending method to obtain the SCR; the aperiodic trend is assumed to correspond to the SCL, and the residual correspond to the SCR [3].

After dismissing activities with missing skin conductance or heart rate data in five participants due to incorrect sensor mounting and/or wireless connectivity issues, we were left with data from 15 participants. All the above features were calculated using 90 s windows with an overlap of 80 s. For each participant, the five features were normalized within participants to zero mean and unit standard deviation to remove most of the large across-participant variability in the raw physiological data. Given the wide range of EDA and HR values as well as the observed variation in the response to stressors, using raw values made it difficult to compare data across participants.

3.3 GEE Regression Models

We developed GEE regression models to predict *stress/relax* labels from the collected physiological variables. The ground truths for these three models were the 1) protocol based labels, 2) subjective based labels, and 3) *ReBreathe* labels, respectively. Given respiration data was used for *ReBreathe* relabeling, we did not include respiratory features as a predictor variable; that is, the independent variables in the GEE regression model only included features derived HRV and EDA.

Generalized linear models (GLMs) are a standard method used to fit regression models for data with binary outcomes. They predict the probability of the occurrence of an event by fitting data to a logit function [64]. Generalized estimating equations were developed to extend the GLM to accommodate correlated data [65]. The prediction outcome Y_i of our GEE model is represented as a linear combination of β_i parameters and the predictor HRV and EDA variables X_i:

$$\begin{aligned} \text{Logit}\,(Y_i|X_i) = \beta_o + \beta_1(\text{AVNN}) + \beta_2(\text{pNN25}) + \beta_3(\text{RMSSD}) \\ + \beta_4(\text{HRV} - \text{HF}) + \beta_5(\text{SCR}), \end{aligned} \tag{1}$$

where β_o is the intercept and β_i the set of regression coefficients for the five predictor variables, computed using an iterative approach to solve the set of estimating equations [65].

To test the three short-term stress prediction models, we computed their predictive accuracy on the 15 participants using a Leave-One-Participant-Out cross validation. Namely, the model was trained on 14 out of the 15 participants and then used to estimate the probability $\rho_{stress}$of the data for the remaining participant using Eq. (1). We used the threshold $\rho_{stress} \geq 0.5$ to label each activity as stressful. This process was repeated for each participant.

4 Results

4.1 Physiological Data

The data collected showed an inconsistency in the response elicited in some participants during DB, thus questioning the validity of protocol-based *stress/relax* labels for all participants. Based on the respiratory behavior of the participants during DB, we divided the participants into two groups: group G1 containing nine participants who were unable to

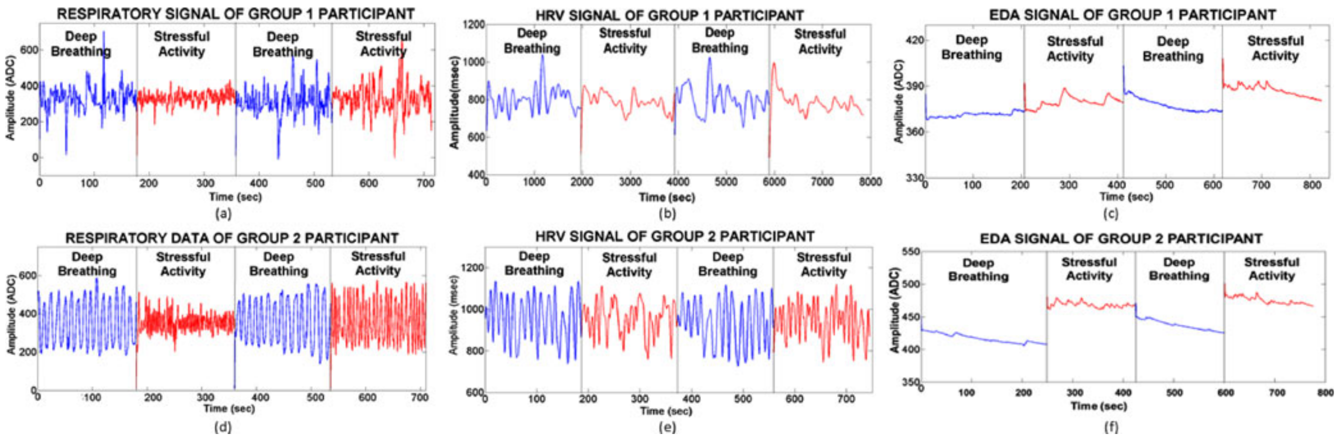

Fig. 3. (a) Respiration, (b) HRV and (c) EDA of one participant in Group 1 during DB (blue) and stress-inducing (red) activities. Plots (d), (e), (f) correspond to a participant in Group 2.

perform DB correctly (defined as having irregular breathing rates during DB), and group G2 whose six participants were able to breathe at the prescribed rate of six breaths per minute for the majority of the DB period. As DB induces relaxation [12], the respiratory signal can be used to predict relaxation [9]. Fig. 3a shows the raw respiratory signal during consecutive DB and stress-inducing activities for a G1 participant in our study. The participant's breathing rate is irregular in both sets of activities (i.e. varies rapidly from shallow to rapid within the acitivity) due to the participant's inability to perform DB at the prescribed rate of 0.1 Hz. In contrast, Fig. 3d shows the respiration signal for a G2 participant who was able to perform DB properly; the respiratory signal is deep and regular during the DB activities due to a decrease in SNS activation as compared to the signal in the stressful activities.

The heart rate and EDA data also shows the same variation between the responses of the two groups during the DB activities. As seen in Fig. 3b, the G1 participant has an irregular heart rate pattern in both sets of activities, indicative of stress even during the DB activities. In contrast, the heart rate of the G2 participant—see Fig. 3e, shows high amplitude and periodicity during the DB activities, indicating the participant was more relaxed. The limited variation in the EDA values of the G1 participant indicates the participant stayed stressed during the DB activities—see Fig. 3c. A comparison of Figs. 3c and 3f shows that the EDA of the G2 participant has less prominent spikes (i.e. SCRs) than that of the G1 participant during the DB activities, indicating a higher level of relaxation. Also, unlike the G1 participant, the EDA values of the G2 participant show a strong decay during both DB activities.

Table 3 presents the mean feature values from each participant. The values clearly illustrate the differences in the responses of the participants in each group. The means of the HRV feature values for G1 are lower (red) whereas the SCR means are higher (green); both of which are indicative of higher stress. The mean SCR and pNN25 of G1 is 5.4 and 0.55, whereas the mean SCR and pNN25 of G2 is 4.3 and 2.58, respectively. The RESP-LF values for G2 are also higher (green) than those of G1. The standard deviations of the feature values are given in Table 4; it shows that G1 participants have the most variation in SCR values whereas G2 participants have the most variation in HRV-HF. Also, the values of G2 vary over a broader range compared to G1.

We used Principal Component Analysis (PCA) to reduce the dimensionality of the feature matrix from six down to two [9] with Fig. 4 showing a scatterplot of the first two principal component scores of the response of the two groups to the DB activities. An independent two-tailed, paired, non-parametric Wilcoxon rank-sum t-test was applied on the first component of the PCA to test the hypothesis that there was no difference in the response of the participants during the DB exercises. The t-test was rejected with $p = 1.6 \times 10^{-17}$ indicating a significant difference between the two groups.

TABLE 3
Mean Feature Values during the DB Activities*

	Mean feature values per Subject (A.U.)						
		SCR	AVNN	pNN25	RMSSD	HRV-HF	RESP-LF
GROUP 1	*S20*	5.57	8.49	0.45	1.21	0.58	4.47
	S16	3.97	6.74	0.26	0.93	0.32	3.77
	S6	7.10	6.89	0.01	1.40	0.71	3.23
	S4	5.76	7.20	0.25	0.95	0.32	3.15
	S11	3.88	7.59	0.71	1.37	0.56	3.06
	S5	6.61	6.79	0.82	1.46	0.92	2.88
	S2	4.54	7.97	0.68	1.27	0.72	2.59
	S1	5.58	7.30	1.23	1.65	0.78	1.90
	S3	5.19	7.29	0.56	1.17	0.48	1.88
	Overall	5.40	7.36	0.55	1.27	0.60	2.99
GROUP 2		SCR	AVNN	pNN25	RMSSD	HRV-HF	RESP-LF
	S19	3.77	7.51	2.49	2.70	2.07	4.49
	S15	4.31	7.76	1.65	2.18	1.80	4.25
	S9	4.54	9.70	4.02	4.64	2.47	4.01
	S14	3.99	9.23	2.83	3.08	4.10	3.65
	S13	5.62	9.17	2.43	2.76	3.05	3.39
	S10	3.76	8.76	2.06	2.41	2.03	3.23
	Overall	4.30	8.69	2.58	2.96	2.59	3.84

**Red – lower values; Green – higher values*

TABLE 4
Standard Deviation of Features during the DB Activities*

	Standard deviation of feature values per Subject (A.U.)						
		SCR	AVNN	pNN25	RMSSD	HRV-HF	RESP-LF
GROUP 1	*S20*	0.78	0.43	0.34	0.25	0.25	1.70
	S16	0.33	0.67	0.13	0.18	0.14	1.28
	S6	0.30	0.39	0.40	0.24	0.31	1.68
	S4	0.57	0.63	0.13	0.19	0.16	1.63
	S11	0.57	0.47	0.57	0.31	0.23	1.53
	S5	1.06	0.21	0.42	0.41	0.63	1.52
	S2	0.24	0.21	0.43	0.31	0.47	0.47
	S1	0.19	0.39	0.80	0.54	0.36	1.35
	S3	0.38	0.11	0.41	0.36	0.46	1.23
	Overall	1.19	0.67	0.51	0.37	0.39	1.52
GROUP 2		SCR	AVNN	pNN25	RMSSD	HRV-HF	RESP-LF
	S19	0.09	0.76	1.56	0.85	1.11	3.06
	S15	0.57	0.31	0.57	0.54	1.21	1.37
	S9	0.14	1.90	1.47	2.19	5.04	1.90
	S14	0.30	0.78	0.49	1.29	4.55	2.66
	S13	0.66	0.31	0.41	0.66	1.03	1.63
	S10	0.68	0.54	0.79	0.48	1.33	2.67
	Overall	0.79	1.17	0.99	1.33	3.33	2.15

**Red – higher values; Green – lower values*

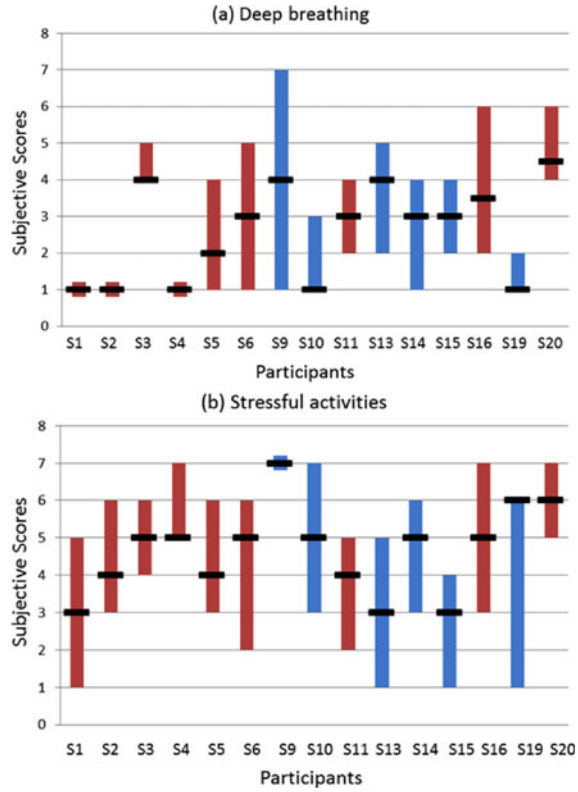

Fig. 5. The range and median of the self-reported stress scores for (a) all six deep breathing activities and (b) all five stress inducing activities given by each of the participants for Group 1 (red) and Group 2 (blue).

4.2 Subjective Scores

Participants provided a rating of stress (1-7 scale) after completion of each activity where 1 corresponded to least stressful, 4 to undecided and 7 to most stressful. The distribution of subjective scores given in Fig. 5a shows an inconsistency in the participants' response to the repeated six DB activities, indicating either difficulties in recall or that their perceived stress levels varied each time they performed the same activity. 13 percent of the scores were 5 and above (i.e. slightly to extremely stressful); 19 percent of the DB activities were scored 4 (undecided) on the Likert scale. The median score of the DB activities was 3. Only five of the participants recorded a score of one (the lowest stress level); of these, only two participants were from G2, a surprising result since G2 represents participants who were able to perform DB. Likewise, nine participants had a median score at or above three; four of these participants were from G2. Participants S1, S2, and S4 rated the DB activities with scores of 1 and S5 with a median score of 2; however, their physiological mean feature values (Table 3) were more indicative of stress. Participant S13 rated the DB activities with a higher median score (4) than the stressful activities (3), S9 also rated the DB activities with median score of 4, however their physiological features values during the DB activities as seen in Table 3 indicates that they were able to DB and relax. Subjective scores for the five stress-inducing activities—see Fig. 5b, also vary over a wide range. 40 percent of the stress activities were scored less than or equal to the midpoint 4 (undecided) and the median score for the stressful activities was 5. However, unlike relaxation activities where the same DB activities were repeated, here participants performed a range of activities.

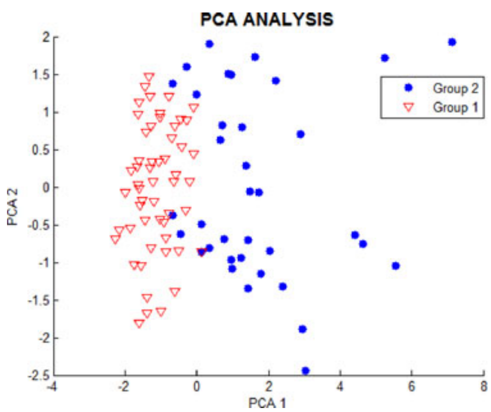

Fig. 4. A scatterplot of the first two principal component scores of the six features for all the participants during the DB activities [9].

The absence of a clear threshold differentiating stress and relax in the subjective scores made it difficult to convert the seven-point Likert subjective scale to a binary *stress/relax* (0/1) label, and specifically how to classify the undecided score of 4. As discussed earlier, this data suggests that some participants felt that the stress activities were not stressful and the DB activities were not relaxing. Given that the median DB score was 3 and the results in Section 4.1 showed that more than half of the participants (G1) were unable to relax in the DB activities, we considered the undecided score of 4 as indicative of stress. We thus labeled all activities that received a subjective score of 4 (undecided) and above as stressful (i.e. greater than the median score of 3) and those below 4 as relaxed.

TABLE 5
Comparison of Original Protocol-Based and New Labels

After relabeling		Subjective-based labels of DB activities			Subjective-based labels of stress-inducing activities			ReBreate labels of DB activities		
		mean	min	max	mean	min	max	mean	min	max
Group 1	**# Labels changed**	2.2	0	6	1.1	0	3	4.8	3	6
	SCR (*relax*)	5.52	4.17	6.91	5.63	5.11	6.18	3.87	2.23	4.55
	SCR (*stress*)	5.33	3.08	7.34	5.82	5.14	7.02	5.60	4.45	7.10
	ΔSCR (%)	-9%	-40%	8%	5%	1%	15%	53%	20%	129%
Group 2	**# Labels changed**	1.7	0	3	2	0	4	1	0	3
	SCR (*relax*)	4.35	3.76	5.74	5.44	4.14	5.96	4.07	3.54	4.54
	SCR (*stress*)	4.68	3.77	6.08	5.22	4.66	5.74	5.20	4.52	5.97
	ΔSCR (%)	1%	-9%	6%	0%	-12%	39%	29%	23%	35%
***p*-value**			***0.56***			***0.66***			***0***	

4.3 Re-Labeling Approaches

In a final step, we compared the three labeling approaches presented in Section 3.1: *ReBreathe* re-labeling, subjective relabeling, and the original protocol labels. To perform respiratory relabeling, we computed the value of RESP-LF (respiratory power in the range 0.04-0.15 Hz) for intervals in the baseline and DB training data when the respiratory signals had a frequency of 0.1 Hz and used it as a threshold, $RESP - LF(threshold)$. Then, we relabeled each DB activity as $y_i = 1$ (stress-inducing) if its RESP-LF value was lower than the threshold, i.e. $RESP - LF(DB_i) < RESP - LF(threshold)$. 54 percent of the 90 DB labels obtained from 15 participants were changed from the original protocol based labels. To obtain the subjective-based labels, we relabeled as $y_i = 1$ (stress-inducing) any activity with $score \geq 4$ (i.e., the mid-point of the Likert scale used). From the original protocol based labels, 33 percent of the DB labels and 34 percent of the stress labels were changed.

To better understand the effect of 1) subjective-based relabeling of DB activities, 2) subjective-based relabeling of stress-inducing activities, and 3) *ReBreathe* relabeling of DB activities, we compared the resulting new labels against the original protocol-based labels. Results are summarized in Table 5. On average, *ReBreathe* relabeling of DB activities for G2 participants (blue) led to the lowest number of changes (1 label) while *ReBreathe* relabeling of the DB activities for G1 participants (red) led to the highest number of changes (4.8 labels). This is consistent with the results in Section 4.1, which showed that the majority of G1 participants were unable to maintain regular breathing rates. For some G1 participants, ReBreathe resulted in changes to all labels (max: 6), suggesting failure to perform DB in any of the six DB activities. On the other hand, for some G2 participants, ReBreathe did not change any of the DB labels (min: 0), indicating participants were able to perform DB correctly. On average, subjective-based relabeling of the DB and stress-inducing activities changed 1-2 labels for both G1 and G2 participants (range: 0-6); this too is consistent with the results presented in Fig. 5.

Given the absence of a reliable ground truth to compare *stress/relax* labels, we also compared the average SCR and change in SCR, ΔSCR = SCR(*stress*)-SCR(*relax*), for G1 and G2 participants (Table 5), as EDA is an independent measure of arousal not affected by respiration [66]. As seen, the average SCR values of subjective-based relabeled *relax* and *stress* activities were similar (G1: $|\Delta SCR| < 10\%$; G2: $|\Delta SCR| \leq 1\%$). For G1 participants, SCR values were similar for relabeled DB (*relax*: 5.52; *stress*: 5.33) and stress-inducing (*relax*: 5.63; *stress*: 5.82) activities, which indicates that their arousal levels were not consistent with their subjective *stress/relax* ratings. Both groups had participants with lower SCR in activities relabeled as *stress* using subjective-based relabeling (G1: min $\Delta SCR = -40\%$; G2: min $\Delta SCR = -12\%$). These inconsistent results support our concern that subjective-based relabeling can be unreliable.

In contrast, *ReBreathe* resulted in a significantly higher ΔSCR for G1 and G2 participants, 53 and 27 percent, respectively. Morever, the average SCR value in activities relabeled as *stress* was higher than activities relabeled as *relax* for both G1 (*stress*: 5.6; *relax*: 3.87) and G2 (*stress*: 5.2; *relax*: 4.07), which indicates that participants were less aroused during *relax* relabeled activities than *stress* relabeled activities. We conducted a two-sided t-test to test the null hypothesis that the average SCR of activities with resulting *relax* labels did not differ from that of activities with *stress* labels in all three categories. Only *ReBreathe* relabeling showed a significant difference at the 0.01 level.

GEE models were then built to predict *stress/relax* labels from the collected HRV/SCR features and validated through a Leave-One-Participant-Out approach. The resulting beta coefficients, and the predictive values (p-values) for the three models are presented in Table 6, show that pNN25 has the largest predictive value. All HRV coefficients have negative polarity, as expected, given that HRV decreases with increasing stress levels. Likewise, the SCR coefficient has positive polarity, as EDA increases with stress. In a

TABLE 6
GEE Regression Coefficients for the Three Models

Feature	Protocol based		Subjective based		*ReBreathe* based	
	coeff	p-value	coeff	p-value	coeff	p-value
AVNN	−0.0023	0.0017	−0.0023	0.0016	−0.0027	0.0014
PNN25	−0.4972	0.0007	−0.8183	0.0007	−1.483	0.0001
RMSSD	−0.0171	0.0048	−0.0525	0.0025	−0.6015	0.0013
HRV HF	−0.0123	0.0024	−0.0013	0.0041	−0.0042	0.0059
SCR	0.0477	0.0047	0.0376	0.0052	0.0014	0.0004

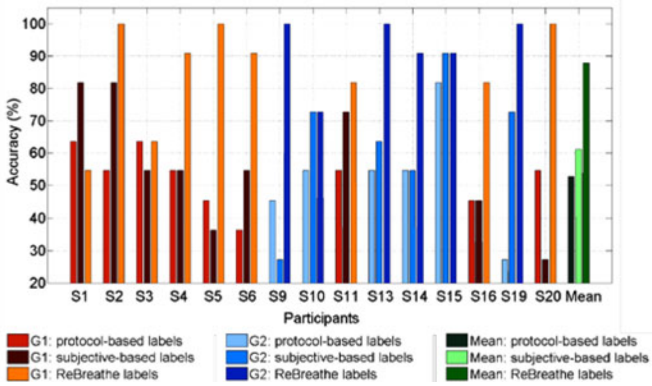

Fig. 6. Classification rate of leave-one-participant-out cross validation of GEE model using protocol-based, subjective-based and ReBreathe labels for model training.

two-sided t-test (H_0: *coefficient* $= 0$), all five coefficients for all three models were significant at the 0.01 level.

The resulting accuracies of the GEE models are summarized in Fig. 6. The protocol-based labels yield an average classification rate of 53 percent. Relabeling the DB activities based on the subjective-based labels increased the mean classification of the resultant GEE model to 61 percent. Using RESP- LF power to re-label the data resulted in a significant[1] increase in the accuracy of the model to a rate of 88 percent. In all but two cases (participants 1, 3), *ReBreathe* relabeling improved stress prediction as compared to using protocol labels, in many cases bringing prediction rates from chance levels (~50 percent) to near-perfect classification (~100 percent). In contrast, using subjective scores for re-labeling yielded inconsistent results across all participants. For four participants, the predictive accuracy fell to even below chance level.

As respiration rate influences HRV [67] but does not affect EDA [66], we also rebuilt the GEE model to predict the *ReBreathe* labels using only SCR. The resultant model accuracy dropped down to 72 percent, which, though lower than the accuracy of the HRV + SCR GEE model, was still greater than the accuracy of models built using the protocol-based and subjective-based labels.

5 Discussion and Conclusions

5.1 Discussion of Results

We analyzed the physiological and subjective response of participants to a series of deep breathing and stress-inducing activities to validate the ground truth *stress/relax* labels of activities used in training stress monitoring devices. We found several contradictions between the participants' responses and standard protocol-based labels, substantiating our concern that activities that induce stress or relaxation may in some cases elicit the opposite response. As seen in Table 3, not all the participants were able to relax during the DB activities. This observation indicates that protocol-based labels for DB relaxation activities may not be applicable across all participants as the ability of DB to elicit relaxation varies from participant to participant. As DB is an acquired skill, not all participants (G1) were able to perform it correctly, which could be due to the short duration of the DB activities in the protocol and/or absence of respiratory biofeedback. Given the correspondence between DB and stress state (i.e., *stress/relax*), the results also show that respiratory data can be used to determine if participants are able to perform DB properly and, as a result, achieve relaxation. This difference in participant response to DB is also reflected in their resulting *ReBreathe* labels; more than half of the DB activities were relabeled as *stress* with some participants staying stressed throughout the protocol. The SCR values (arousal level) for DB activities relabeled as *stress* was significantly different to that of DB activities relabeled as *relax* across all participants (Table 5).

The subjective scores (Fig. 5) provided by the participants were also not consistent with their physiological data as presented in Table 3. These inconsistencies indicate that perceived stress scores do not necessarily correspond to the physiological changes; possibly due to the participants' unfamiliarity with deep breathing. This lack of consistency makes subjective scores unreliable markers to qualify individual stress perception in our protocol. The results in Table 5 support these inconsistencies, where there was minimal difference in arousal levels (SCR) between subjective-based relabeled *stress* and *relax* activities. Neither the subjective labels obtained from participants' perceived scores nor the protocol-based labels agreed with the resultant *ReBreathe* labels. Similarly, some of the participants did not perceive the stress-inducing activities as being stressful, providing subjective scores as low as 1, possibly due to their unfamiliarity with these activities. Though self-reported stress levels are often used to infer the psychological impact of stress on individuals, these results show that due to their subjective nature, they cannot be generalized both within and across participants.

Results from the physiological measurements and subjective scores put into question the assumptions used in standard protocol-based labeling (i.e., that DB elicits a state of relaxation for all participants) and subjective-based labeling (i.e., self-reporting accurately qualifies individual stress perception). Instead, the high variability within and between individuals suggests that protocol-based and subjective-score based labels cannot be used to train stress detection models, and support the need for relabeling. The accuracy of GEE models trained using labels obtained with *ReBreathe* was 35 and 27 percent better than accuracies obtained with models developed using standard nominal labels and subjective stress scores, respectively. As respiration rate influences HRV [67], relabeling with respiration improves the accuracy of the stress prediction model. EDA, on the other hand, provides an independent measure of arousal and has been shown to only co-vary with respiration, not controlled by respiration [66]. Our results indicate that the labels assigned to DB activities need to be validated to ensure participants are relaxed and not stressed, before they can be used to train an accurate stress prediction model.

Inducing relaxation in individuals is more difficult than inducing stress [11]. It is possible that the short

1. A two-tailed, paired, non-parametric Wilcoxon rank-sum t-test was performed to compare the respiration based model to the protocol and subjective score based models (H_0: the two models have the same error rate). The hypothesis was rejected in both cases with a p-value $<$ 0.05.

duration of the DB activities between the stressful activities in our protocol was not a sufficient period for relaxation to be induced in some of the participants. Most guidelines and studies looking at the impact of DB ask users to perform these exercises for 10 minutes and longer over multiple sessions (e.g., [11], [37], [40]). Furthermore, the absence of respiratory biofeedback during the DB exercises could have also led to some participants not being able to estimate their respiratory rate to correctly deep breathe and therefore relax. Respiratory relabeling was not applied to stress-inducing activities since participants did not control their breathing rates during these activities. Nevertheless, the varied subjective scores provided by participants for the stress-inducing activities in our study indicate that relabeling stress activities similarly should be explored.

5.2 Conclusions

As the goal of stress management is to prevent the onset of chronic stress, devices are needed to detect physiological warning signs of stress that can harm health. Though significant work has been done in developing devices for stress monitoring [4], [6], accurately training these devices remains an open problem. The standard approach is to either use protocol labels [4], [6] or perceived stress scores [46], [49] to label activities in the training dataset. However, given the high rate of variability within and amongst participants in their response to stress, these approaches do not provide an accurate ground truth about the individual response. Using accurate activity labels based on a user's individual response to stress and relaxation can result in improvements in the accuracy of the stress prediction model. Given the inaccuracies in protocol-based and subjective based labeling as seen in our results; a physiological measure (respiration) with known links to relaxation offers an accurate calibration method for these devices.

In this study, we proposed a relabeling method, *ReBreathe*, which uses the respiratory signal to determine the accurate *stress/relax* labels of DB exercises in a calibration protocol. *ReBreathe* takes advantage of the fact that DB should effect respiration to determine baseline relaxation labels. The low frequency component of the respiratory signal can thus be used to identify if the participant is breathing at the prescribed rate and actually able to relax to determine the participant's true label for that activity. As seen in Section 4.3, it is possible to accurately predict stress labels obtained from the calibration respiratory data utilizing only HRV and EDA as well as only EDA features. *ReBreathe* can similarly be used to train a relaxation system providing users with feedback about the efficacy of their breathing technique via heart rate and skin conductance measures only. *ReBreathe* can also be used to normalize the high level of inter-participant variability observed in the perceived stress scores for DB exercises. Correlating self-reported scores to DB activities with validated relax labels allows normalization of the DB scores to be performed so they map uniformly within and across all individuals.

5.3 Further Work

Our study was conducted on a relatively small sample size ($n = 15$ participants). Additional work is required to validate *ReBreathe* on a larger population. We also need to determine whether longer DB periods might be more efficient in inducing relaxation and stress activities should also be similarly relabeled. One of our immediate targets is to validate *ReBreathe* and our stress prediction model in ambulatory settings. In that context, we plan to utilize our earlier work on the removal of motion artifacts from physiological signals [68] to calibrate the device for when the participant is mobile. Calibrating individual subjective stress scores could allow subjective stress ratings to be included as an additional input in stress prediction models to provide a multilevel stress prediction output instead of a binary output.

Acknowledgments

The authors would like to thank the associate editor and anonymous reviewers, whose comments have helped improve this paper. The authors also thank the volunteers for their participation in the data recording. The work was supported by NPRP grants # [08-125-2-031] and # [5-678-2282] from the Qatar National Research Fund which is a member of Qatar Foundation. The statements made herein are solely the responsibility of the authors.

References

[1] H. Selye, *The Stress of Life*. New York, NY, USA: McGraw Hill, 1978.

[2] J. D. Clark, D. R. Rager, and J. P. Calpin, "Animal well-being ii. stress and distress," *Comparative Med.*, vol. 47, no. 6, pp. 571–579, 1997.

[3] J. Choi, B. Ahmed, and R. Gutierrez-Osuna, "Development and evaluation of an ambulatory stress monitor based on wearable sensors," *IEEE Trans. Inf. Technol. BioMed.*, vol. 16, no. 2, pp. 279–286, Feb. 2012.

[4] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Troster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 1, pp. 410–417, Mar. 2010.

[5] J. Wijsman, B. Grundlehner, H. Liu, H. Hermens, and J. Penders, "Towards mental stress detection using wearable physiological sensors," in *Proc. IEEE Ann. Int. Conf. Eng. Med. and Biol. Soc.*, Aug. 2011, pp. 1798–1801.

[6] A. Barreto, J. Zhai, and M. Adjouadi, "Non-intrusive physiological monitoring for automated stress detection in human-computer interaction," in *Proc. IEEE Int. Conf. Human-Comput. Interaction*, 2007, pp. 29–38.

[7] M. Kusserow, O. Amft, and G. Tröster, "Monitoring stress arousal in the wild," *IEEE Pervasive Comput.*, vol. 12, no. 2, pp. 28–37, Apr.-Jun. 2013.

[8] K. Richardson and H. Rothstein, "Effects of occupational stress management intervention programs: A meta-analysis," *J. Occupational Health Psychol.*, vol. 13, pp. 69–93, 2008.

[9] H. Khan, B. Ahmed, J. Choi, and R. Gutierrez-Osuna, "Using an ambulatory stress monitoring device to identify relaxation due to untrained deep breathing," in *Proc. IEEE Int. Conf. Eng. Med. & Bio. Soc.*, Osaka, Japan, 2013, pp. 1744–1747.

[10] K. Masood, B. Ahmed, J. Choi, and R. Gutierrez-Osuna, "Consistency and validity of self-reporting scores in stress measurement surveys," in *Proc. IEEE Int. Conf. Eng. Med. Bio. Soc.*, San Diego, CA, USA, 2012, pp. 4895–4898.

[11] H. Benson, *The Relaxation Response*. New York, NY, USA: Harper Torch, 2000.

[12] M. Nogawa, T. Yamakoshi, A. Ikarashi, S. Tanaka, and K. Yamakoshi, "Assessment of slow-breathing relaxation technique in acute stressful tasks using a multipurpose non-invasive beat-by-beat cardiovascular monitoring system," in *Proc. IEEE Int. Conf. Eng. Med. & Bio. Soc.*, 2007, pp. 5323–5325.

[13] D. Martarelli, M. Cocchioni, S. Scuri, and P. Pompei, "Diaphragmatic breathing reduces postprandial oxidative stress," *J. Alternative Complementary Med.*, vol. 17, no. 7, pp. 623–628, 2011.

[14] R. Rosmond, “Role of stress in the pathogenesis of the metabolic syndrome,” *Psychoneuroendocrinology*, vol. 30, pp. 1–10, 2005.
[15] (2011). Mental health: facing the challenges, building solution, world health organization [Online]. Available: http:// www.euro.who.int/__data/assets/pdf_file/0008/96452/E87301.pdf
[16] Stress related and psychological disorders in Great Britain, health and safety executive [Online]. Available: http://www.hse.gov.uk/statistics/causdis/stress/stress.pdf, 2014
[17] (2010). Psychologically healthy workplace program - fact sheet by the numbers, american psychological association [Online]. Available: http://www.apa.org/practice/programs/ workplace/phwp-fact-sheet.pdf
[18] [Online]. Available: http://www.stress.org/what-is-stress/, 2015
[19] R. Mermelstein, S. Cohen, and T. Kamerck, “A global measure of perceived stress,” *J. Health Social Behavior*, vol. 24, pp. 385–396, Dec. 1983.
[20] L. Crockett, J. Schulenberg, and A. Petersen, “Congruence between objective and self-report data in a sample of young adolescents,” *J. Adolescent Res.*, vol. 2, pp. 383–392, 1987.
[21] J. Pruessner, D. Hellhammer, and C. Kirschbaum, “Burnout, Perceived Stress, and Cortisol Responses to Awakening,” *Psychosomatic Med.*, vol. 61, pp. 197–204, 1999.
[22] S. Cohen, D. Tyrrell, and A. Smith, “Negative life events, perceived stress, negative affect, and susceptibility to common cold,” *J. Personality Social Psychol.*, vol. 64, no. 1, pp. 131–140, 1993.
[23] J. Smyth, M. Ockenfels, L. Porter, C. Kirschbaum, D. Hellhammer, and A. Stone, “Stressors and mood measured on a momentary basis are associated with salivary cortisol secretion,” *Psychoneuroendocrinology*, vol. 23, no. 4, pp. 353–370, 1998.
[24] R. Orsila, M. Virtanen, T. Luukkaala, and M. Tarvainen, “Perceived mental stress and reactions in heart rate variability–a pilot study among employees of an electronics company,” *Int. J. Occupational Safety and Ergonom.*, vol. 14, no. 3, pp. 275–283, 2008.
[25] S. Bakewell, “The autonomic nervous system,” vol. 5, 1995.
[26] F. Martini and E. Bartholomew, *Essentials of Anatomy & Physiology*. San Francisco, CA, USA: Benjamin Cummings, 2003, p. 267.
[27] A. Ben-Tal, S. Shamailov, and J. Paton, “Central regulation of heart rate and the appearance of respiratory sinus arrhythmia: New insights from mathematical modeling,” *Math. Biosci.*, vol. 255, pp. 71–82, 2014.
[28] J. H. Houtveen, S. Rietveld, and E. J. Geus, “Contribution of tonic vagal modulation of heart rate, central respiratory drive, respiratory depth, and respiratory frequency to respiratory sinus arrhythmia during mental stress and physical exercise,” *Psychophysiology*, vol. 39, no. 4, pp. 427–436, 2002.
[29] L. Freedman, A. Scerbo, M. Dawson, A. Raine, W. McClure, and P. Venables, “The relationship of sweat gland count to electrodermal activity,” *Psychophysiology*, vol. 31, no. 2, pp. 196–200, Mar. 1994.
[30] P. Grossman, “Respiration, stress, and cardiovascular function,” *Psychophysiology*, vol. 20, no. 3, pp. 284–300, 1983.
[31] C. Wientjes, “Respiration in psychophysiology: Methods and applications,” *Biological Psychol.*, vol. 34, no. 2–3, pp. 179–203, 1992.
[32] E. J. deGeus, G. H. Willemsen, C. H. Klaver, and L. J. V. Doornen, “Ambulatory measurement of respiratory sinus arrhythmia and respiration rate,” *Biological Psychol.*, vol. 41, no. 3, pp. 205–227, 1995.
[33] L. Varvogli and C. Darviri, “Stress management techniques: Evidence-based procedures that reduce stress and promote health,” *Health Sci. J.*, vol. 5, no. 2, pp. 74–89, 2011.
[34] C. S. Disorders, “Relaxation techniques,” 2011.
[35] J. F. Thayer, S. S. Yamamoto, and J. F. Brosschot, “The relationship of autonomic imbalance, heart rate variability & cardiovascular disease risk factors,” *Int. J. Cardiol.*, vol. 141, no. 2, pp. 122–131, 2010.
[36] R. Jerath, J. Edry, V. Barnes, and V. Jerath, “Physiology of long pranayamic breathing: Neural respiratory elements provide a mechanism that explains how slow deep breathing shifts the autonomic nervous system,” *Med. Hypotheses*, vol. 67, pp. 566–571, 2006.
[37] G. Pal, S. Velkumary, and M. Trakoo, “Effect of short-term practice of breathing exercises on autonomic functions in normal human volunteers,” *Indian J. Med. Res.*, vol. 120, no. 2, pp. 115–121, 2004.
[38] S. Z. Wang, S. Li, X. Y. Xu, G. P. Lin, L. Shao, Y. Zhao, and H. W. Ting, “Effect of slow abdominal breathing combined with biofeedback on blood pressure and heart rate variability in prehypertension,” *J. Alternative Complementary Med.*, vol. 16, no. 10, pp. 1039–1045, 2010.
[39] S. M. Stewart, C. Betson, T. H. Lam, I. B. Marshall, P. W. H. Lee, and C. M. Wong, “Predicting stress in first year medical students: a longitudinal study,” *Med. Educ.*, vol. 31, pp. 163–168, 1997.
[40] G. Paul, B. Elam, and S. Verhulst, “A longitudinal study of students’ perceptions of using deep breathing meditation to reduce testing stresses,” *Teaching Learning Med.*, vol. 19, no. 3, pp. 287–292, 1997.
[41] E. Forbes and R. Pekala, “Psychophysiological effects of several stress management techniques,” *Psychological Rep.*, vol. 72, pp. 19–27, 1993.
[42] H. Toivanen, E. Lansirnles, V. Jokela, and O. Hanninen, “Impact of regular relaxation training on the cardiac autonomic nervous system of hospital cleaners and bank employees,” *Scandanvian J. Work Environ. Health*, vol. 19, pp. 319–325, 1993.
[43] D. Martarelli, M. Cocchioni, S. Scuri, and P. Pompei, “Diaphragmatic breathing reduces exercise-induced oxidative stress,” *Evidence-Based Complementary Alternative Med.*, vol. 2011, pp. 1–10, 2011.
[44] “Wild Divine,” Wild Divine, Inc. [Online]. Available: http://www.wilddivine.com/, 2015
[45] HeartMath, Emwave Personal Stress Reliever [Online]. Available: http://www.myemwave.org, 2015
[46] J. Parkka, J. Merilahti, E. Mattila, E. Malm, K. Antila, M. Tuomisto, A. V. Saarinen, M. van Gils, and I. Korhonen, “Relationship of psychological and physiological variables in long-term self-monitored data during work ability rehabilitation program,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 2 pp. 141–151, Mar. 2009.
[47] J. Ogorevc, A. Podlesek, G. Gersak, and J. Drnovsek, “The effect of mental stress on psychophysiological parameters,” in *Proc. IEEE Int. Workshop Med. Meas. and Appl.*, 2011, pp. 294–299.
[48] A. de santos Sierra, C. Avila, J. Guerra Casanova, and G. del Pozo, “A stress-detection system based on physiological signals and fuzzy logic,” *IEEE Trans. Ind. Electron*, vol. 58, no. 10, pp. 4857-4865, Oct. 2011.
[49] K. Plarre, A. Raij, S. Hossain, A. Ali, M. Nakajima, M. Al’absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, D. Siewiorek, A. Smailagic, and L. Wittmers, “Continuous inference of psychological stress from sensory measurements collected in the natural environment,” in *Proc. 10th Int. Conf. Inform. Process. Sensor Netw.*, 2011, pp. 97–108.
[50] D. Reeves, S. LaCour, K. Winter, K. Raynsford, and K. Vogel, “The UTC-PAB/AGARD STRES Battery: User’s manual and system documentation,” *Storming Media*, 1991.
[51] [Online]. Available: http://psych.hanover.edu/JavaTest/CLE/Cognition/Cognition/dualtask_instructions.html, 2013
[52] J. L. Edens, K. T. Larkin, and J. L. Abe, “The effect of social support and physical touch on cardiovascular reactions to mental stress,” *J. Psychosomatic Res.*, vol. 36, no. 4, pp. 371–381, 1992.
[53] J. Stroop, “Studies of interference in serial verbal reactions,” *J. Exp. Psychol.*, vol. 18, pp. 643–62, 1935.
[54] C. Kirschbaum, K. M. Pirke, and D. H. Hellhammer, “The ‘trier social stress test: A tool for investigating psychobiological stress in a laboratory setting,” *Neuropsychobiology*, vol. 28, no. 1–2, pp. 76–81, 1993.
[55] S. Cohen, T. Kamarck, and R. Mermelstein, “A global measure of perceived stress,” *J. Health Social Behavior*, vol. 24, pp. 385–396, 1983.
[56] S. Cohen and J.-D. D., “Who’s stressed? distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009,” *J. Appl. Social Psychol.*, vol. 42, no. 6, pp. 1320–1334, 2012.
[57] K. Trimmel, “Sensitivity of HRV parameters including pnnxx proven by short-term exposure to 2700m altitude,” *Physiological Meas.*, vol. 32, no. 3, pp. 275–85, 2011.
[58] P. Stein and R. Kleiger, “Insights from the study of heart rate variability,” *Ann. Rev. Med.*, vol. 50, pp. 249–61, 1999.
[59] M. Dawson, A. Schell, and D. Filion, “The electrodermal system,” in *Handbook of Psychophysiology*, 3rd ed., Cambridge, U.K.: Cambridge Univ. Press, 2007, pp. 159–81.

[60] J. Mietus, C. Peng, I. Henry, R. Goldsmith, and A. Goldberger, "The pNNx files: Re-examining a widely used heart rate variability measure," *Heart*, vol. 88, pp. 378–80, 2002.
[61] M. Malik, J. Bigger, A. Camm, R. Kleiger, and et al., "Heart rate variability standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, pp. 1043–1065, 1996.
[62] D. Clifton, J. G. Douglas, P. S. Addison, and J. N. Watson, "Measurement of respiratory rate from the photoplethysmogram in chest clinic patients," *J. Clinical Monitoring Comput.*, vol. 21, pp. 55–61, 2007.
[63] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *J. Neurosci. Methods*, vol. 190, pp. 80–91, 2010.
[64] A. Agresti and B. Finlay, "Logistic regression: Modeling categorical responses," in *Statistical Methods for the Social Sciences*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1997, pp. 688–724.
[65] S. L. Zeger and K.-Y. Liang, "Longitudinal data analysis for discrete and continuous outcomes," *Biometrics*, vol. 42, pp. 121–130, 1986.
[66] R. Schneider, S. Schmidt, M. Binder, F. Schaffer, and H. Walach, "Respiration-related artifacts in EDA recordings: Introducing a standardized method to overcome multiple interpretations," *Psychological Reports*, vol. 93, no. 3, pp. 907–920, 2013.
[67] J. D. Schipke, G. Arnold, and M. Pelzer, "Effect of respiration rate on short-term heart rate variability," *J. Clinical Basic Cardiol.*, vol. 2, no. 1, pp. 92–95, 1999.
[68] F. Alamudun, J. Choi, R. Gutierrez-Osuna, H. Khan, and B. Ahmed, "Removal of subject-dependent and activity-dependent variation in physiological measures of stress," in *Proc. 6th Int. Conf. Pervasive Comput. Technol. Healthcare*, San Diego, CA, USA, 2012, pp. 115–122.

Beena Ahmed received the BEng degree from the University of Engineering and Technology in Lahore, Pakistan in 1993 and the PhD degree from the University of New South Wales, Sydney, Australia in 2004, both in electrical engineering. She is currently a senior lecturer at the Texas A&M University at Qatar, Doha, Qatar. She was a postdoctoral fellow with the RMIT University, Melbourne, Australia, till 2005. Her research interests include biomedical signal processing, physiological stress modeling, speech processing, feature extraction, noise reduction, EEG processing, and sleep analysis. She is a member of the IEEE.

Hira Mujeeb Khan received the BS degree in telecommunication engineering from National University of Computer and Emerging Sciences, Islamabad, Pakistan, in 2010. She was working as research assistant at Texas A&M University at Qatar, Doha, Qatar. Her research interests include affective computing, psychophysiology, and wireless networks. She is a member of the IEEE.

Jongyoon Choi received the BS degree in industrial engineering from the Sunfkyunkwan University, Seoul, Korea, in 1995, the MS degree in comuter engineering from Gwangju Institute of Science and Technology, Gwangju, Korea, in 1998, and the PhD degree in computer science from Texas A&M University, College Station, in 2011. He is currently working with the Intel Corporation at Austin, Texas. His research interests include pattern recognition, affective computing, wearable sensors, and psychophysiology.

Ricardo Gutierrez-Osuna received the BS degree in electrical engineering from the Polytechnic University, Madrid, Spain, in 1992, and MS and PhD degrees in computer engineering from North Carolina State University, Raleigh, in 1995 and 1998, respectively. He is currently a professor of computer engineering at Texas A&M University, College Station. His current research interests include wearable physiological sensors, voice and accent conversion, speech and face perception, and active chemical sensing biography appears here. He is a senior member of the IEEE.