Expert Systems With Applications 140 (2020) 112890

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Analysis of multimodal physiological signals within and between individuals to predict psychological challenge vs. threat

Aya Khalaf[a], Mohsen Nabian[b], Miaolin Fan[c], Yu Yin[b], Jolie Wormwood[d], Erika Siegel[e], Karen S. Quigley[d], Lisa Feldman Barrett[d], Murat Akcakaya[a], Chun-An Chou[c], Sarah Ostadabbas[b,*]

[a] Electrical & Computer Engineering Department, University of Pittsburgh, PA, USA
[b] Electrical & Computer Engineering Department, Northeastern University, MA, USA
[c] Mechanical & Industrial Engineering Department, Northeastern University, MA, USA
[d] Department of Psychology, University of New Hampshire, NH, USA
[e] Department of Psychology, University of California San Francisco, CA, USA

ARTICLE INFO

Article history:
Received 27 January 2019
Revised 30 July 2019
Accepted 18 August 2019
Available online 19 August 2019

Keywords:
Challenge vs. threat
Multimodal data fusion
Mutual information
Support vector machines
Clustering
Graph theory

ABSTRACT

Challenge and threat characterize distinct patterns of physiological response to a motivated performance task where the response patterns vary as a function of an individual's evaluation of task demands relative to his/her available resources to cope with the demands. Challenge and threat responses during motivated performance have been used to understand psychological, behavioral, and biological phenomena across many motivated performance domains. In this study, we aimed to investigate individual and group-level variations in physiological responding across a series of motivated performance tasks that vary in difficulty. The proposed approach is motivated by documented individual differences in physiological responses observed in motivated performance tasks, such that we first focus on individual differences in physiological responses rather than group-level comparisons. Then, through our analysis of individuals we identify sub-groups (i.e., clusters) of individuals that share common physiological patterns across tasks of varying difficulty and we perform across-subject analysis within each cluster. This is distinct from existing studies which typically do not examine individual vs. subgroup-specific patterns of physiological activity. Such an approach enables us to identify patterns in physiological responses that can be used to predict self-reported judgments of challenge vs. threat with higher accuracy in each subgroup compared to an analysis that includes the entire sample population as a single group. Specifically, three hypotheses were tested: (H1) individuals will have different sets of physiological patterns (features) across tasks of varying difficulty; (H2) there will be subgroups of individuals who share common salient physiological features across the subgroup clusters that differentiate their physiological responding across tasks of varying difficulty; and (H3) the accuracy of predicting self-reported judgments of challenge vs. threat across individuals will be higher within each subgroup with shared salient physiological features than across all subgroups or the entire sample with all computed features. To test these hypotheses, we developed an integrated analytic framework for multimodal physiological data analysis. We employed data from an existing experiment in which participants completed three mental arithmetic tasks of increasing difficulty during which different modalities of physiological data were collected. Analyses revealed three subgroups of participants who shared common features that best differentiated their within-individual physiological response patterns across tasks. Support vector machine (SVM) classifiers were trained using both shared features within each group and all computed features to predict challenge vs. threat states. Results showed that, the within-group classification model using group common features achieved higher self-report prediction accuracy compared to an alternative model trained on data from all participants without feature selection.

* Corresponding author.
E-mail addresses: afk17@pitt.edu (A. Khalaf), monabiyan@ece.neu.edu (M. Nabian), fan.mi@husky.neu.edu (M. Fan), yinyu@ece.neu.edu (Y. Yin), Erika.Siegel@ucsf.edu (E. Siegel), K.Quigley@northeastern.edu (K.S. Quigley), l.barrett@northeastern.edu (L.F. Barrett), akcakaya@pitt.edu (M. Akcakaya), ch.chou@northeastern.edu (C.-A. Chou), Ostadabbas@ece.neu.edu (S. Ostadabbas).

https://doi.org/10.1016/j.eswa.2019.112890
0957-4174/

1. Introduction

Psychologists have documented distinct patterns of psychological and cardiovascular response during motivated performance situations—situations that are goal-relevant to the performer, require instrumental cognitive responses, and are active rather than passive (e.g., public speaking, sports competitions, test taking; Blascovich & Mendes, 2000). A growing body of research has specifically focused on distinctions between two biopsychological response patterns, challenge and threat, which are related to an individual's perception of situational and task demands (including perceptions of danger, uncertainty, and required effort) relative to perceptions of one's available resources to cope with these demands (Blascovich & Mendes, 2000; Tomaka, Blascovich, Kelsey, & Leitten, 1993; Tomaka, Blascovich, Kibler, & Ernst, 1997). Challenge is experienced when a person judges his/her coping resources to meet or exceed the demands of the situation or task, whereas threat is experienced when a person judges the situation or task demands to exceed his or her coping resources (Blascovich & Mendes, 2000; Tomaka et al., 1993, 1997). Challenge and threat states are associated with important outcomes, like task performance. For example, when compared to individuals experiencing challenge, individuals experiencing threat performed worse on the math portion of the Graduate Record Examinations (GRE) which is a test that measures quantitative reasoning, critical thinking, analytical writing, and verbal reasoning skills (Jamieson, Mendes, Blackstock, & Schmader, 2010). Challenge and threat states are also commonly observed during potentially stressful interpersonal interactions. Individuals are more likely to experience threat when interacting with individuals who are higher in social status (Mendes, Blascovich, Major, & Seery, 2001), are from a racial group other than their own (Mendes, Blascovich, Lickel, & Hunter, 2002; Blascovich, Mendes, Hunter, Lickel, & Kowai-Bell, 2001; Berry Mendes, Gray, Mendoza-Denton, Major, & Epel, 2007), are socially stigmatized (such as having a visible facial birthmark; Blascovich et al., 2001), or when their interaction partner violates a known stereotype (such as speaking with an Asian male with a southern accent; Mendes, Blascovich, Hunter, Lickel, & Jost, 2007).

According to the biopsychosocial model (e.g., Blascovich & Mendes, 2000; Blascovich & Tomaka, 1996), challenge is associated with enhanced ventricular contractility (as measured by pre-ejection period or PEP) and increased cardiac output (CO), as well as decreased systemic vascular resistance (measured as total peripheral resistance or TPR) (Blascovich & Mendes, 2000; Dienstbier, 1989), a pattern of cardiovascular reactivity that should increase blood flow to skeletal muscles and the heart, thereby supporting greater motor activity. Threat, on the other hand, is associated with no change or increases in systemic vascular resistance (Blascovich & Mendes, 2000; Dienstbier, 1989), a pattern of cardiovascular reactivity that reduces efficient and effective blood flow to peripheral blood vessels, and thereby provides poor or limited support of increased muscle action. In the psychophysiological literature (Tomaka et al., 1993, 1997; Mendes, Major, McCoy, & Blascovich, 2008; Jamieson, Nock, & Mendes, 2012; Quigley, Barrett, & Weinstein, 2002), challenge and threat have typically been indexed via patterns of task-related change in pre-ejection period (PEP), heart rate (HR or the inverse of HR, inter-beat interval or IBI), cardiac output (CO) and/or stroke volume (SV), and total peripheral resistance (TPR).

In the decades since their scientific introduction, challenge and threat have been used to understand psychological, behavioral, and biological responses across many motivated performance domains, particularly those involving social evaluation (Mendes et al., 2001, 2002; Blascovich et al., 2001; Berry Mendes et al., 2007; Mendes et al., 2007). However, the vast majority of this work has not examined challenge and threat at the individual level, focusing instead on group comparisons between samples of individuals who either exhibit more threat-like versus more challenge-like subjective experiences or more threat-like or challenge-like changes in cardiovascular physiology. A recent meta-analysis on the existing empirical literature using this group-level approach found stable relationships between cardiovascular physiology and performance, but these relationships were fairly weak (Behnke and Kaczmarek, 2018), suggesting there is still room for improvement in predicting experience and behavior from physiological activity during motivated performance contexts. A renewed focus on within-individual comparisons may allow for the identification of potentially important individual differences in physiological response patterns that could better predict experience and behavior across varying motivated performance contexts.

In one exception to the commonly-used group-based approach, Quigley et al. (2002) investigated individual patterns of cardiovascular responding across a series of motivated performance tasks of increasing difficulty. In this experiment, participants completed four mental arithmetic tasks, which required them to perform serial subtractions aloud (e.g., "subtract sevens from the number 3,746") in the presence of an evaluator. Each iteration of the task increased in difficulty by adding new elements of social evaluative pressure as the experiment progressed. The experience of psychological challenge and threat (as measured by self-reported stress and coping resources) were assessed before each task and cardiovascular physiological reactivity was measured before, during and after every task. Analyses revealed individual differences across genders in terms of both cardiovascular reactivity and appraisals of stress and coping.

These findings provide a preliminary demonstration of the importance of examining individual-specific patterns of psychological and physiological responding during motivated performance, as opposed to responses at the group-level. Adopting an individual-focused approach is also consistent with the idea that challenge and threat are malleable states that exist as opposing endpoints of a continuum (as opposed to fixed, dichotomous states; Seery, 2013; Jamieson et al., 2016). The appraisals of demands and resources associated with challenge and threat are posited to occur on a more subconscious or automatic level and to change dynamically over time as perceived demands or resources shift (see, Seery, 2013; Quigley et al., 2002). Given this, examining *changes* in patterns of physiological activity over time within individuals may provide a more nuanced means of predicting experience and behavior during motivated performance contexts than static, aggregated measures taken at the group level.

1.1. Our contributions

In the present study, we intend to extend the existing literature on physiological responding during motivated performance in several ways. First, although previous research has investigated both within- and between-subject variability in responding during affective events more broadly (Wu & Parsons, 2011; Bozhkov, Georgieva, Santos, Pereira, & Silva, 2015; Kim & André, 2008; Chen et al., 2016), these innovations have not yet been leveraged in the literature on physiological responding during motivated performance tasks where there is a scarcity of research examining variability in response patterns. The vast majority of the literature utilizing the biopsychosocial model, for example, has examined group-level differences in two specific physiological patterns (challenge vs. threat) within a single motivated performance context across stipulated groups (e.g., groups based on subject variables like race or gender, or based on an experimenter-defined condition). Here, to examine individual differences, we adopted an approach similar to Quigley et al. (2002) in which participants completed a series of motivated performance tasks of increasing difficulty. Then, changes in peripheral physiological responding across tasks within-individuals were examined–instead of examining a mean or aggregate level of physiological responding across individuals, as is more typically done. We hypothesized that we would observe more individually-diverse patterns of peripheral physiological response during motivated performance than what has been previously recognized.

Second, we examined whether additional physiological variables could help distinguish individual differences in self-reported responding to motivated performance tasks of increasing difficulty. Recent studies examining physiological responding during motivated performance from the biopsychosocial perspective have focused almost exclusively on cardiovascular responses during motivated performance, without systematically investigating the possible utility of additional indices of peripheral physiological activity for understanding biological responding in active coping stress tasks. However, there are potential advantages to using additional multimodal information so that prediction accuracy can be improved (for example, as in Chen, Tao, Huang, Miyasato, & Nakatsu, 1998), and potentially also examine whether physiological variables beyond cardiovascular ones are useful in distinguishing stress and coping experiences. To address this, in the present study, we included a much broader range of peripheral physiological measures than is typical in previous studies on challenge and threat, and used them as input to the proposed feature selection and classification algorithms. In addition to indices of cardiovascular responding that are typically investigated within this literature, we also examined measures of facial muscle activity (including activity in the corrugator supercilia and zygomaticus major muscle groups), respiration, and electrodermal activity. A schematic of the proposed study design, data pre-processing and data analysis is shown in Fig. 1.

In the present investigation, we first focused on examining within-individual variation in physiological activity during a series of motivated performance tasks of increasing difficulty (i.e., with increasing task demands over iterations of the task). It was hypothesized that, within individuals, patterns of physiological responding would change as the task demands increased (Hypothesis 1). Then, person-specific physiological features that best differentiated within-individual physiological responding across tasks of varying difficulty were identified. Based on the existing literature from the biopsychosocial perspective, one might assume that changes in cardiovascular reactivity (i.e., PEP, IBI, CO, and TPR) would best distinguish physiological responding across tasks of varying difficulty for all individuals (as individuals likely move from feeling more challenged to more threatened as task difficulty increases). However,

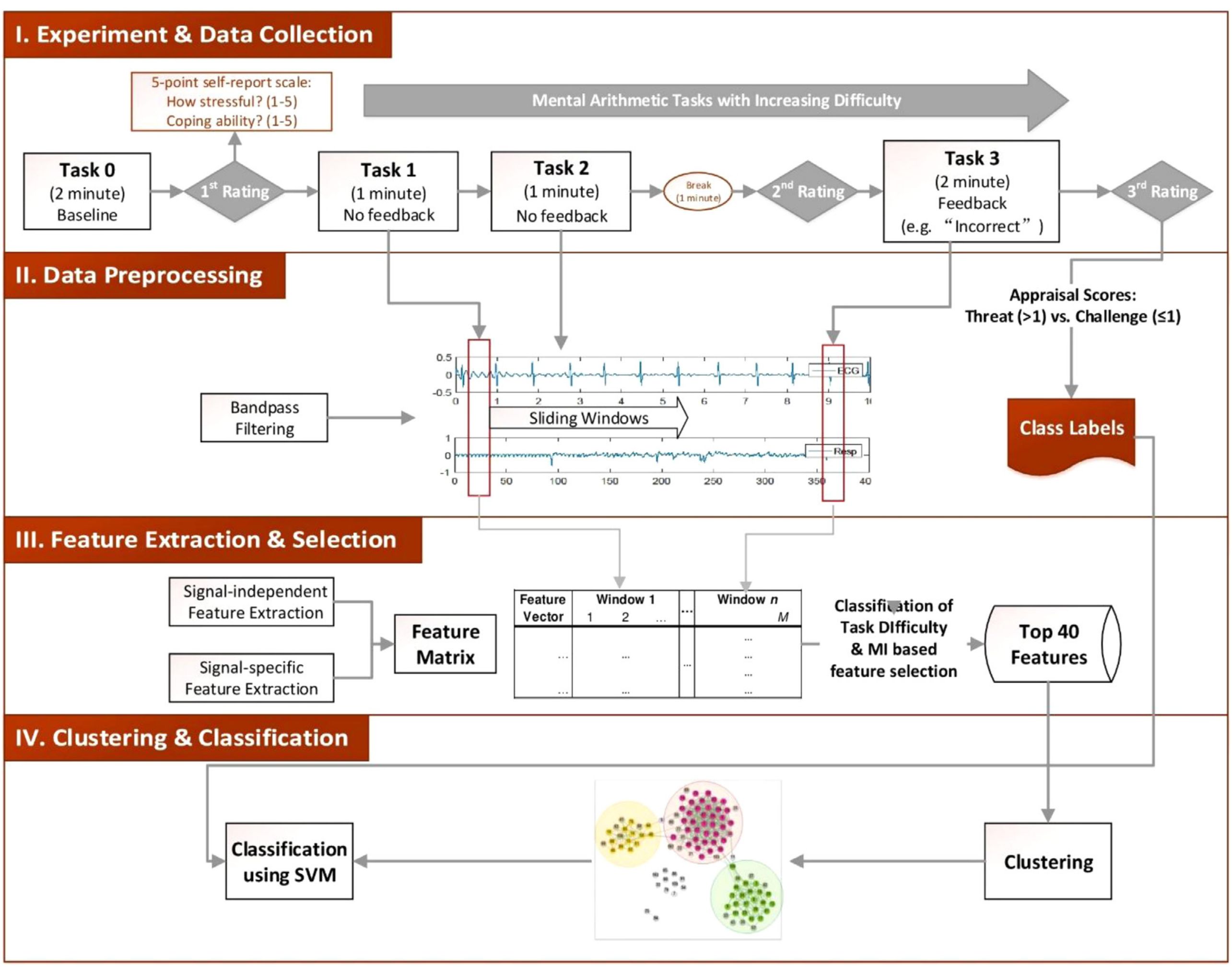

Fig. 1. The methodological framework of this study: (I) In the experiment phase, four mental arithmetic tasks with increasing difficulty were conducted to collect multimodal physiological data as well as self-report scores from participants. (II) Following a band-pass filter and normalization, sliding windows (10s length for EDA and 5s length for other modalities) with 50% overlap were used for segmenting the signals. (III) Two types of feature extraction approaches were applied to the segmented data to obtain the feature matrix, where each row represented a feature vector of one participant. Then, we selected top 40 features based on the mutual information between feature vectors and class labels (specified based on the task difficulty level). (IV) A distance-based clustering algorithm was performed based on the selected subset of features, and SVM classifier was trained for within and across group classification.

emerging empirical and theoretical evidence suggests that mental states, including emotional and motivational states, do not appear to have consistent and specific physiological 'fingerprints' in the body across all individuals, but rather variability across individuals appears to be the norm (for a discussion, see Siegel et al., 2018). Thus, here it was hypothesized that there would be different groups (clusters) of participants, each of which share common salient features that dominate this differentiation in physiological responding across tasks (Hypothesis 2). That is, it was predicted that not all participants' physiological responses across tasks would be best differentiated by the same small subset of cardiovascular variables. A network analysis approach was then applied to characterize these groups and investigate how the shared salient features varied across groups. Finally, we hypothesized that post-task self-reports of stress and perceived coping resources would be better predicted using the shared salient features within each group versus using the same set of physiological features for all individuals in the sample as a whole (Hypothesis 3). By leveraging a data set with a larger set of peripheral physiological variables than are typically used in the literature on motivated performance, coupled with an individual difference-focused analytic strategy, we hypothesized that we would reveal heretofore undetected patterns of physiological change across active coping stress tasks of varying difficulty that could then be used to more accurately predict stress experience during motivated performance.

In summary, our novel contributions in this work are twofold: (1) examining the individualized differences in multimodal physiological responses during motivated performance tasks with varying difficulties; and (2) identifying clusters of individuals that share common physiological responses and performing group-level analysis of predicting self-reported judgments of challenge vs. threat within each cluster. Such an approach enables us to identify patterns in physiological responses to predict self-reported judgments of motivated performance tasks with higher accuracy in each cluster compared to an analysis that includes the entire sample population in a single group.

2. Materials and methods

2.1. Participants

260 participants were recruited from the greater Boston area to participate in this experiment. Here, data from the 107 participants (41 males, 65 females; 1 gender not reported) who had complete, clean physiological data for all tasks and all minutes of the baseline were used. The remaining 152 participants had data from at least one physiological measure for at least 1 min of either the tasks or baseline that was not usable (due to excessive movement artifact, noisy signal caused by improper skin preparation or sensor placement, etc.). Participants were required to be native English speakers and were excluded if: they were currently taking any medications that would influence peripheral physiological activity, had a history of mental illness or cardiovascular disease, or reported consuming caffeine, alcohol, or tobacco within 24 h prior to the experiment. All participants completed the mental arithmetic tasks as part of a single 3-4-h experimental session in which they completed a series of other tasks and questionnaires not relevant to the current investigation. The study was approved by the Institutional Review Board at Northeastern University.

2.2. Measures

2.2.1. Physiological responding signals

Electrocardiogram (ECG), impedance cardiogram (ICG), continuous blood pressure (BP), respiration (RES), electrodermal activity (EDA), and measures of facial muscle activity over the corrugator supercilii (COR) and zygomaticus (ZYG) major muscle groups using facial electromyography (fEMG) were recorded. All physiological measures were sampled at 1000 Hz using BioLab v. 3.0.13 (Mindware Technologies; Gahanna, OH) via a BioNex 8-Slot chassis (Model 50-3711-08). Samples of filtered physiological data from one participant are shown in Fig. 2. The filter settings for each biosignal is later discussed in the Data Preprocessing section.

ECG was obtained using pre-gelled Ag/AgCl sensors in a modified lead II configuration. ICG was acquired using a four-spot electrode configuration (see Qu, Zhang, Webster, and Tompkins, 1986) using pre-gelled Ag/AgCl electrodes. The inner (recording) electrodes were placed on the participant's chest: one at the base of the neck at the top of the sternum and the other at the level of the xiphisternal junction. The outer (source) electrodes were placed along the midline on the participant's back approximately 4 cm above and below the inner recording electrodes (respectively, roughly over the fourth cervical vertebrae and the ninth thoracic vertebrae). The source electrodes passed a 4 mA, 100 kHz alternating current across the thorax. BP was recorded continuously via a Continuous Noninvasive Arterial Pressure monitor (CNAP Monitor 500AT; CNSystems; Medizintechnik, AG, Austria). Continuous recordings were obtained from small cuffs placed on participants' left middle and pointer fingers, and these continuous readings were calibrated intermittently against an automated non-invasive, blood pressure cuff around the participant's right upper arm. The RES signal was measured via a piezoelectric belt placed around the lower-chest/upper-abdomen (Mindware Technologies; Model 50-4504-00). The EDA signal was recorded from the palmar surface of the right hand with sensors on the thenar and hypothenar eminences using disposable, Ag/AgCl (11 mm diameter; isotonic paste) electrodes (Biopac Systems, Inc.; Goleta, CA). fEMG measures were obtained via reusable Ag/AgCl electrodes (Mindware Technologies; Gahanna, OH) filled with an electrolyte gel and placed over the zygomaticus major and corrugator supercilii muscle regions on the right side of the participant's face. A reference electrode was placed in the middle of the forehead.

2.2.2. Procedure: mental arithmetic tasks

Each participant completed a series of mental arithmetic tasks as part of a single experimental session (see Appendix A for the description of the full experimental session), which typically lasted between 3 and 4 h, during which s/he completed a number of other tasks unrelated to the current investigation. Following electrode placement, the participant was seated in a sound-attenuated testing room in an upholstered chair. Immediately before completing the series of mental math tasks pertinent to the current investigation, the participant was connected to the continuous blood pressure monitor, and the continuous readings from the finger cuffs on the left hand were calibrated against a non-invasive reading taken from the cuff on the right arm. The participant then sat quietly alone in the testing room while resting physiological measures were recorded for 2 min (*Task 0* or Baseline). The participant was not informed prior to this baseline that s/he would be doing arithmetic tasks immediately following the baseline. This was done to ensure that this resting baseline was not confounded with anticipatory stress related to the upcoming tasks.

After the baseline, the participant was informed that s/he would complete a series of mental arithmetic tasks by speaking serial subtraction answers aloud in front of an experimenter, who would record the answers. The participant was asked to work as quickly and accurately as possible, and to refrain from commenting on the task or his or her performance until all tasks were complete. The participant then practiced by subtracting aloud from 90 by 3 s for 30 s and was encouraged to ask any questions about the

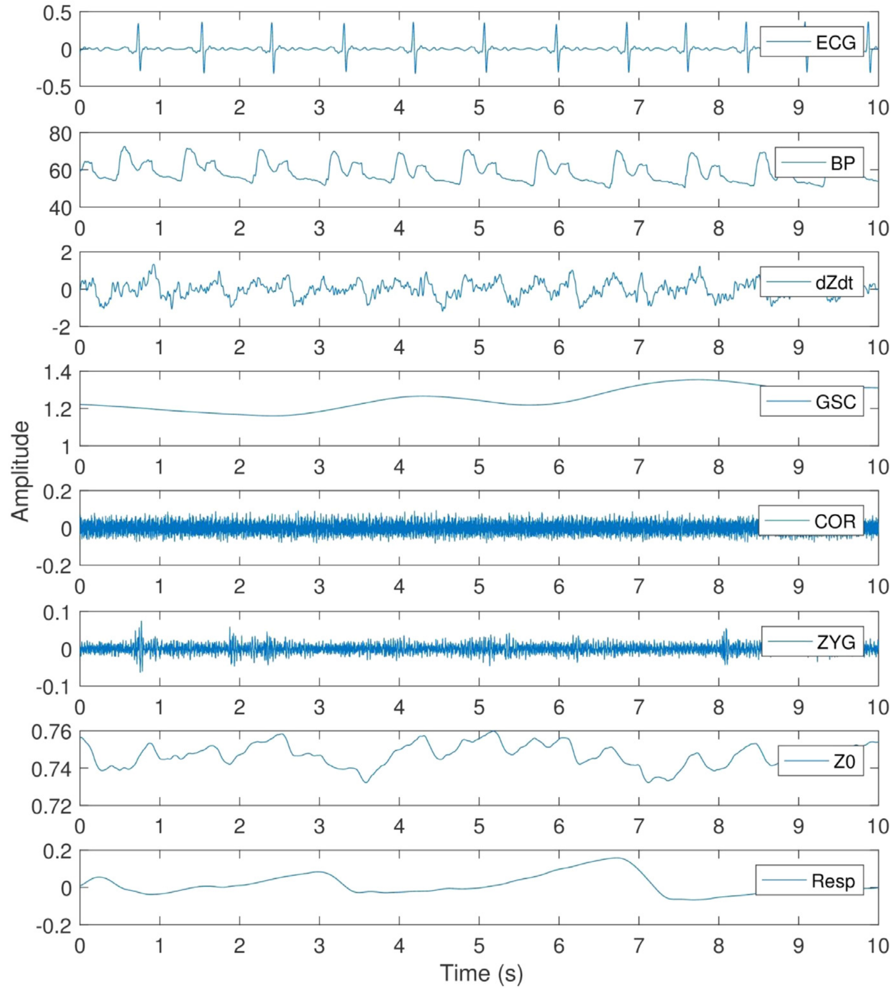

Fig. 2. Samples of collected physiological signals after preprocessing.

task instructions. Next, the participant rated how stressful the upcoming task would be and how well s/he could cope with the upcoming task on two 5-point scales (from 1 = Not at All to 5 = Very Much). Immediately following these ratings, the first subtraction task began and the participant was asked to subtract from 725 by 7 s for 1 min while an experimenter recorded the responses (*Task 1*). Experimenters were trained not to provide any positive feedback (e.g., no smiling or nodding) during the task. For the second arithmetic task, the participant again subtracted aloud for 1 min, this time from 847 by 6 s, 8 s, or 13 s with the goal of keeping the serial subtractions moderately stressful for all participants by adjusting the task difficulty to their initial performance (*Task 2*). Again, the experimenter recorded his or her responses, but gave no feedback.

The participant next was given a short (1 min) break to sit quietly alone in the testing room, after which the experimenter returned to the room, and the participant was told that there would be one last math task, but that this time s/he would subtract from a larger number and that the experimenter would inform the participant each time an incorrect response was given. These changes were meant to increase the stressfulness of the serial subtractions in the final 2-min task. Before these subtractions, the participant again rated how stressful s/he thought the upcoming task would be and how well s/he could cope with the upcoming task on the same 5-point scales. The participant then was asked to subtract from 4851 by 8 s, 12 s, or 17 s for 2 min (again depending on performance during the first mental arithmetic task) (*Task 3*). For these 2 min, the experimenter recorded the participant's answers and provided feedback each time s/he was incorrect (e.g., "Incorrect. 4826"). Immediately following this final 2-min task, the participant rated how stressful the task was and how well s/he coped during the task using the same 5-point scales. The ratio of the two ratings (stress rating/coping rating) is defined as the participant's self-reported experience of threat or challenge (*SR*). Ratios greater than "1" indicate a threat experience, while self-report ratings of less than or equal to "1" indicate a challenge experience.

2.3. Hypothesis-driven models

This section describes the methodologies that are used to test the above listed three hypotheses. For the first hypothesis, we aim to formulate a classification problem to distinguish among 4 classes/tasks and our aim is to identify subset of person-specific features (salient patterns from the recorded measurements) which contribute significantly to the classification/separation across tasks of varying difficulties. Then for Hypothesis 2, based on the person-specific features as identified in Hypothesis 1, we perform a clustering based on graphical approach to identify sub-groups of individuals who share person-specific physiological features. Finally for Hypothesis 3, based on the sub-groups identified in Hypothesis 2, we perform an analysis to investigate how the classification accuracy to distinguish between self-reported ratings of threat and challenge experience varies within and across sub-groups.

2.3.1. Problem formulation to test H1

With the aim of finding the physiological features that best differentiate a given individual's pattern of bodily response across tasks of varying difficulty, a classifier will be trained to differentiate 4 classes (T_0 = baseline, T_1 = Task 1, T_2 = Task 2, T_3 = Task 3) for each participant. These features are also believed to be the most experience-indicative features. For a given individual, *i*, let's extract *M* features from the available physiological signals collectively to form feature vector $F^i = [f_1^i,\ f_1^i, \dots, f_M^i]$ in each class of $\{T_0, T_1, T_2, T_3\}$. The objective is to find a subset of K_i feature from F^i in which the classification accuracy when differentiating these 4 classes is more than a predefined accuracy threshold, $\mathcal{T}$.

2.3.2. Problem formulation to test H2

To test this hypothesis, the objective is to find *J* clusters (groups) of individuals, $\{C_1, \dots, C_J\}$ that share a common subset of salient features in differentiating $\{T_0, T_1, T_2, T_3\}$ classes, with a predefined commonality measure greater than $\mathcal{L}$. More specifically, each participant is represented by a binary feature vector of size $F \times 1$ where *F* is the total feature size. In this binary vector if the *i*th feature is significantly contributing to the classification of the 4 tasks/classes then a 1 is placed in the *i*th location; otherwise a 0 is placed. To cluster the participants based on the similarity in the binary feature vectors, a graph was constructed by considering each participant as a node and the Jaccard coefficient (Jaccard, 1912) was considered as the connectivity measure between the binary vectors. A connection is decided to exist between two nodes if the normalized Jaccard coefficient between these nodes is above the predefined commonality measure C = 0.5.

2.3.3. Problem formulation to test H3

To test this hypothesis, an inference model, $\mathfrak{T}$ is defined to predict self-report ratings reported at the end of Task 3 using the feature vector extracted from class T_3. The self-report rating *SR* is predicted in a binary fashion: $SR > 1$ for threat and $SR \leq 1$ for challenge. To predict the self-report ratings *SR* of the participants in the cluster $C_j, j\ \varepsilon [1, \dots, J]$, $\mathfrak{T}$ will be trained in the following structures and tested using a leave-one-out paradigm:

(i) Training $\mathfrak{T}_i$ on the union of the K_i feature subset of each participant in the cluster C_j and testing it across the participants in the cluster C_j.
(ii) Training $\mathfrak{T}_{ii}$ on the union of the K_i feature subset of each participant in the cluster C_j and testing it with the participants outside the cluster C_j.
(iii) Training $\mathfrak{T}_{iii}$ on the union of the K_i feature subset of each participant outside the cluster C_j and testing it with the participants in the cluster C_j.
(iv) Training $\mathfrak{T}_{iv}$ with all the features on all participants of C_k, where $k\ \varepsilon [1, \dots, J]$ and testing the model with the participants from cluster C_j.
(iv) (v) Training $\mathfrak{T}_v$ on all feature set *F* and across all of the *N* test participants, without any clustering and testing it with all of the participants.

2.4. Physiological data preprocessing

In this section we explain the signal processing approaches used for pre-processing of the recorded physiological signals for the replicability of the proposed analysis. Specifically, we describe the filters used to remove noise and extract portions of the data that are of interest. Physiological signals including Electrocardiogram (ECG), Electrodermal Activity (EDA) and Electromyogram (EMG) were processed using the Biosignal-Specific Processing (Bio-SP) Tool developed at Augmented Cognition Lab (ACLab) at Northeastern University (Biosignal-Specific Processing (Bio-SP) Tool, 2018). The Bio-SP Tool extracts features on these signals based on the state-of-the-art studies reported in scientific literature (Perez-Rosero, Rezaei, Akcakaya, & Ostadabbas, 2016; Pan & Tompkins, 1985; Bsoul, Minn, Nourani, Gupta, & Tamil, 2010; De Chazal et al., 2003; K. H. Kim, Bang, & Kim, 2004; Benedek & Hazlett, 2005; Tkach, Huang, & Kuiken, 2010). Processing of other signals such as Blood Pressure (BP), Respiration (RES) and Impedance Cardiography (ICG) were also performed in MATLAB based on the recommendations suggested in scientific articles (Tomaka et al., 1993, 1997; Young, 2001; Allen et al., 1990). In this section, the pre-processing steps are explained, and in the following sections, the data segmentation and feature extraction of the signals are elaborated.

An elliptic band-pass filter with cut-off frequencies of 5–45 Hz was used for pre-processing of ECG signals. These cut-off frequency values were selected based on the power spectral density (PSD) analysis of the ECG signals. In addition, the elliptic filter was selected to ensure that the amplitude of the signal peak points were not significantly suppressed by the filter (Chavan, Agarwala, & Uplane, 2005).

For the EDA signals, FIR low-pass filter with 0.5 Hz cutoff frequency was used (Figner & Murphy, 2011). An elliptic bandpass filter with cutoff frequencies of 10–300 Hz was used for both the COR and ZYG signals based on inspection of the power spectral density of these signals (De Luca, Gilmore, Kuznetsov, & Roy, 2010). Moreover, according to the power spectrum of the two fEMG signals acquired for each participant, two notch filters at 60 Hz (COR) and 180 Hz (ZYG) were used.

For continuous BP in adult humans, the recommended low pass cut-off frequency is in the range of 100–200 Hz (Young, 2001). Based on the power spectral density analysis of the BP signals, a cut-off frequency of 100 Hz was selected for the present sample. The respiration signal (RES) and the transthoracic basal impedance signal (Z0) were filtered with a Butterworth low pass filter with a 20 Hz cut-off frequency.

Finally, for the dz/dt signal derived from ICG, a 2nd order Butterworth bandpass filter with cut-off frequencies of 0.75–40 Hz was used based on the recommendation provided in BioLab software (Knowledge Base Impedance Cardiography).

2.5. Physiological data segmentation

This section describes how the filtered recorded physiological signals are segmented before feature extraction. Data segments corresponding to Task 1 (T_1), Task 2 (T_2), and Task 3 (T_3) as well as the baseline (T_0) were obtained using time stamps provided in a separate text file along with the physiological data file of each participant. In order to extract features from the segment corresponding to each task, a 5-s sliding window was employed with 50 % overlap for 7 out of the recorded 8 signals. Since the EDA is a slowly changing signal, a sliding window of 10 s duration and 50 % overlap was used. Each window was considered as an observation belonging to a specific task and used for feature extraction. The same window size and overlap were used for both within and between individual classification problems. T_0 reduce variability due to the differences in signal range from person to person in the between-subject problem, the features extracted from each baseline window were subtracted from the features of the corresponding T_3 window.

2.6. Feature extraction

This sections describes the types of (signal-specific and signal-independent) features (salient patterns/attributes) extracted from each segmented physiological recording. Several sets of features were extracted from the 8 recorded physiological signals and employed in a machine learning framework to initially differentiate patterns of physiological responding across tasks of varying difficulty levels. The goal is to find the set of features that is most related to within-person changes in bodily responding across various task demands, as such a set of features should correspond with changes in subjective experience across tasks. The features included signal-independent and signal-specific features. Signal-independent features included Fourier based features while signal-specific features included morphological features designed for extracting information from the ECG, fEMG, EDA, BP, and ICG (dz/dt and z0) signals. No morphological features were extracted from RES signal since applying morphology analysis on such signal will result in drastic decrease in the number of data windows that can be obtained from each task. In particular, the task duration varies from 30 to 60 s and RES signal requires around 25-s window for the analysis to obtain representative morphological features for that window whereas the other signals require much smaller window and, therefore, will obtain higher number of data windows for which features can be extracted.

2.6.1. Signal-independent features

The power spectrum was estimated for each window of data using Welch method (Welch, 1967). The raw power spectrum values were considered as features. For each window per signal, the number of features obtained from the power spectrum was reduced by calculating the average power over a narrow range of frequencies instead of using all the values as features. The average of power spectrum values within a sliding window of specific width was considered as one feature. To calculate the same feature for the next window, the original window was shifted by a value equal to its width such that there was no overlap between consecutive windows and the average of that window was calculated and so on. For ECG, RES, and ICG (dz/dt and Z0) signals, the average power was calculated using a window of 2.5 Hz width. Since the BP signal has higher bandwidth compared to the previously stated signals and considering the need to reduce the number of features, a window of 5 Hz width was used. For the same reason, a window of 10 Hz width was employed for the fEMG signals (COR and ZYG).

2.6.2. Signal-specific features

ECG features: In this paper, within each window, we calculated several measures associated with heart rate variability from the ECG signal, including mean R-R intervals (i.e., the time between consecutive heartbeats), standard deviation of R-R intervals, standard deviation of the differences between adjacent R-R intervals, and the square root of the mean of the sum of the squares of differences between adjacent R-R intervals. Another set of calculated features included the number of pairs of adjacent R-R intervals where the first R-R interval exceeded the second R-R interval by more than 50 ms as well as the number of pairs of adjacent R-R intervals where the second R-R interval exceeded the first R-R interval by more than 50 ms. Mean area of each QRS complex was also calculated in addition to its standard deviation.

fEMG features: fEMG records the electrical activity produced by skeletal muscles in the face. The calculated features include the standard deviation, root mean square, and mean absolute value of the fEMG signal within each window. In addition, features related to the signal frequency were calculated, including the number of zero crossings and the number of times the slope sign changed. Another feature was the cumulative length of the fEMG signal within the analysis window, which provides a measure of the complexity of the signal. In addition, an estimate of the exerted muscle force (f) for a signal x of N samples (Ekman, 1993; Tkach, Huang, & Kuiken, 2010) was calculated as follows:

$$f = e^{\frac{1}{N}\sum_{k=1}^{k=N} \log x_k} \tag{1}$$

EDA features: Electrodermal activity (EDA) measures changes in the skin's electrical conductivity due to changes in the amount of sweat present in the eccrine sweat glands of the palm. The EDA signal is composed of two activities, tonic activity represented by the slowly varying base signal and phasic activity or skin conductance responses (SCRs) which are represented by faster variations in the signal. SCRs were detected by performing differentiation and subsequent convolution with a 20-point Bartlett window. In this paper, we extracted the signal mean (mean skin conductance or MSC), the number of detected SCRs, the mean SCR duration, the mean SCR amplitude, and the mean SCR rise-time (where rise-time of an SCR is defined as the time between the initial rise and the peak of an SCR).

BP features: The characteristic points of the BP waveform within each heartbeat were used for calculating 5 signal-specific features. In particular, diastolic pressure defined as the minimum signal value of each heartbeat and systolic pressure defined as the maximum signal value of each heartbeat were considered as features. Pressure values at both the dicrotic notch and the notch peak were also calculated. Mean arterial pressure (MAP), which represents mean signal value for each heartbeat, was also calculated as a feature.

ICG features: Features calculated from the Impedance Cardiogram included left ventricular ejection time (LVET), stroke volume (SV), cardiac output (CO), pre-ejection period (PEP) and total peripheral resistance (TPR). LVET is the time period in which blood flows across the aortic valve, and PEP is the time period between the electrical activity signaling the start of ventricular contraction and the onset of blood being ejecting into the aorta (Goertz, 1995). SV is defined as the blood volume pumped by the heart per heartbeat, and CO is the blood volume pumped by the heart per minute (Maceira, Prasad, Khan, & Pennell, 2006). Finally, TPR represents the resistance of the systemic circulation that carries oxygenated blood from the heart to the body and returns deoxygenated blood to the heart. These 5 variables were extracted from the shape characteristics of ICG (dz/dt) signal combined with the corresponding ECG signal, ICG (Z0) signal, and BP signal. The methods used to extract these features are based upon the discussions provided in the scientific literature (Allen et al., 1990; Kelsey & Guethlein, 1990).

2.7. Feature selection

This section describes the proposed methodology for selection person-specific features for the classification of 4 tasks with varying difficulties, please see also Section 2.3.1 for Hypothesis 1 formulation. In general, the objectives of feature selection are three-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying process that generated the data. In this work, feature selection is related to the within-individuals classification of bodily responses across the 4 tasks of varying difficulty. Feature selection methods include filter, wrapper, and embedded methods. Compared to wrapper and embedded methods, the advantage of using filter methods is the low computational complexity (Saeys, Inza, & Larrañaga, 2007). However, filter methods assume that the features are independent. Therefore, it is possible that these methods select redundant features. Due to the high dimensionality of the feature vector (159 features), a filter method was applied for feature selection. Filter methods include statistical hypothesis tests such as student-t test and Wilcoxon Rank Sum test, however, these tests were not used in this study since they can work only with binary classification problems. In this work, mutual information, which is a filter method, was used to select the most significant K_i features out of the original feature vector $F^i = [f_1^i,\ f_1^i, \ldots, f_M^i]$ for participant *i*. Mutual information measures the contribution of each feature to make a correct decision and assigns each feature a score based on that contribution. In other words, the higher the score is, the higher the contribution is of that feature to correct classification. In this paper, the scores were normalized by the maximum score which corresponds to the most significant feature.

2.8. Task and self-report rating classification

This section describes the classifier used in the proposed methodology for within subject classification of 4 tasks with varying difficulties (Hypothesis 1), and within and across subgroups self-report rating classification. In this paper, for both within-subject task level classification, and across-subjects self-report rating classification, a support vector machine (SVM) classifier with a linear kernel function was used to reduce computational expenses. For across-subjects classification, the goal of SVM is to find an optimal hyperplane that discriminates individuals in terms of self-report ratings ($SR > 1$ versus $SR \leq 1$) in the selected feature space. The classification/prediction accuracy is defined as the percentage of individuals correctly classified in the class they supposedly belong to.

3. Results

3.1. Testing Hypothesis 1: a within-individual analysis

For each participant, a four-class classification problem has been formulated to distinguish among Baseline, Task 1, Task 2, and Task 3. Due to the limited number of samples per task, a leave-one-out cross validation procedure was employed. In this classification problem, mutual information was used to select the most relevant {1, 5, 10, 20, . . ., 159} features and the corresponding accuracy was calculated. Fig. 3 shows the average accuracy obtained over 107 subjects at different numbers of selected features. It is clear that using at least 40 features results approximately in the same accuracy obtained using the whole feature set. Therefore, for each participant, out of 159 possible features, $K_i = 40$ features are

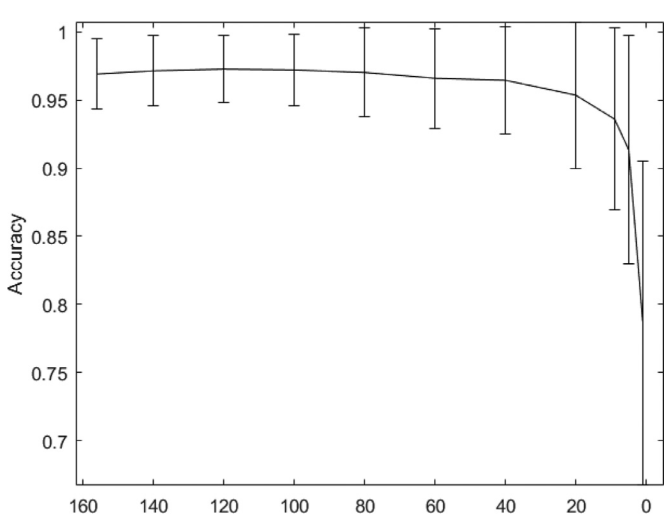

Fig. 3. The average accuracy obtained from 107 subjects plotted at different numbers of selected features.

labelled as the most significant ones. This corresponds to features with normalized mutual information scores higher than 0.7, which also corresponds to a classification accuracy threshold of $\mathcal{T} = 95\%$.

To explore the distribution of the top 40 subject-specific features across individuals, a histogram showing the distribution of the significant features across the 107 subjects included in the study was plotted. The results are illustrated in Fig. 4. This figure shows that among 159 features, the top 20 features appeared across the majority of the individuals included features extracted from EDA, ZYG, and Z0 signals while the top 40 features included same features as well as features extracted from COR and BP signals. Also, note that since the histogram is not uniform, it can be concluded that some features appear in the top 40 feature list more frequently than the others. This also means that different individuals have different significant features in their top 40 list.

3.2. Testing Hypothesis 2: between subject clustering analysis

We hypothesized that there are subgroups of people each of which share common subset of features in their top $K_i = 40$ selected features. Moreover, some individuals may not belong to any group. Finding these subgroups is a clustering problem. However, since the number of the clusters is unknown, and also since we did not want to force any individual to be a part of a certain subgroup/cluster, graph theoretic approach was considered.

As described in Section 2.3.2, a binary feature vector of size 159×1 (the same size as *F*) was formed and for each individual, if the *i*th feature is in his/her top 40 features, we place 1 in the *i*th location of this binary feature vector. Otherwise, a 0 is placed in that location. A connection is decided to exist between two nodes/participants based on the normalized Jaccard coefficient between these nodes.

Fig. 5 shows the clustering patterns of participants in a graph, where three major clusters can be visually observed. The nodes in grey represent individuals that do not belong to any clusters. On the other hand, some of the individuals may have very few connections to one cluster, or some may have connections to multiple clusters. In order to resolve these issues, the cumulative distribution function (CDF) for node degrees of the graph was computed, see Fig. 6. Degree for each node is defined as the number of connections that that node makes with the other nodes in the graph. Based on the degree CDF, a threshold, ρ is identified such that (1) if a node has connections only to one cluster with the connectivity degree of *d*, and if $d > \rho$, this node is determined to belong to that cluster; (2) if a node has connections to more than one cluster (i.e., *J* clusters); assuming that the degree of connectivity to different clusters are defined as $d_1, d_2, \ldots, d_J$, if $max(d_1, d_2, \ldots, d_J) > \rho$, then that belongs to the cluster with the maximum connectivity degree. Here, we selected $\rho = 5$ which corresponds to degree CDF higher than 0.7.

By considering the most frequently observed significant features for the participants within each cluster, shown in Table 1 where SD and SI represent signal-dependent and signal-independent features respectively, it was found that the first cluster containing 15 participants was dominated by the signal-independent features from the COR signal, while the second cluster containing 22 participants was dominated by the signal-independent features derived from the Z0 signal. Moreover, the third cluster with 37 participants was dominated by the ZYG signal-independent features. These observations reveal that, for the majority of individuals in the present study, changes in facial muscle activity were the most common significant features for distinguishing within-person differences in physiological responding across tasks of varying difficulty. For the remaining approximately 1/3 of participants, changes related to cardiac impedance (signal-independent features of the Z0 signal) were the most common significant features, which is more

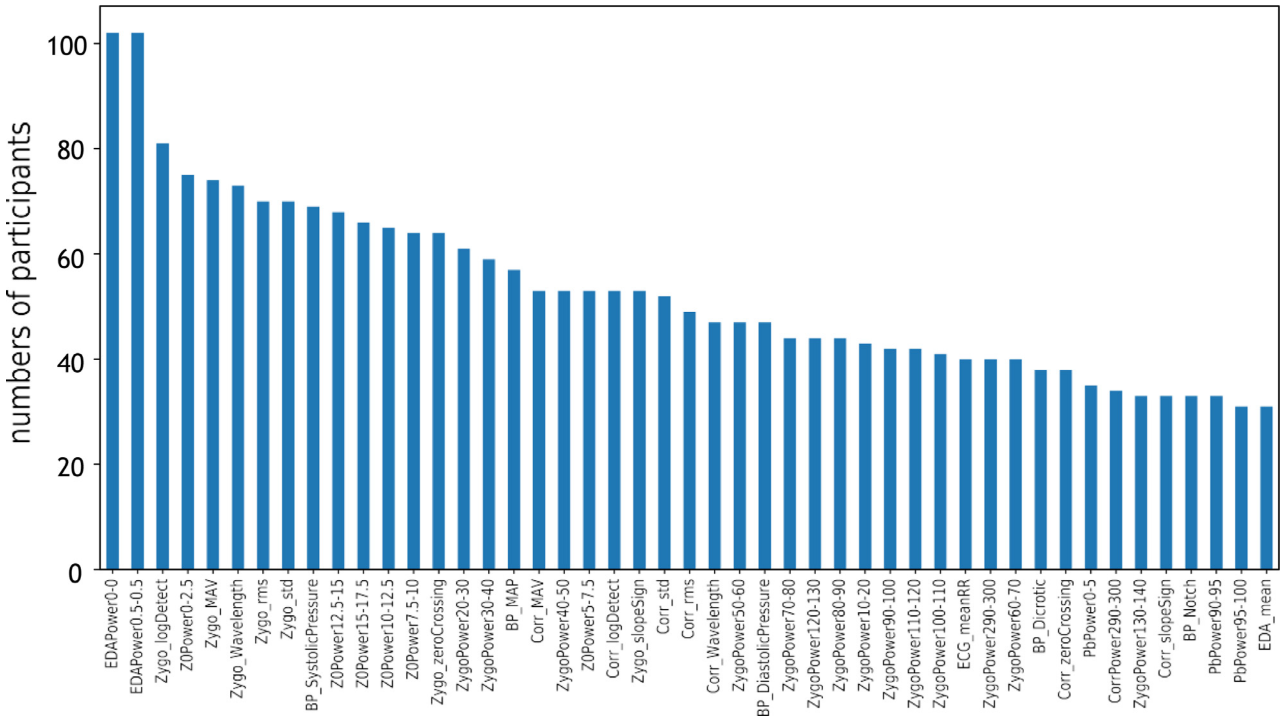

Fig. 4. Distribution of the features appearing among the top Ki = 40 features across all individuals. (only features appeared in at least one-third of all individuals are shown, 47 features out of total 159 features). Features associated with numbers are the power spectrum features. Numbers existing beside these features reflect the frequency range over which each feature is calculated.

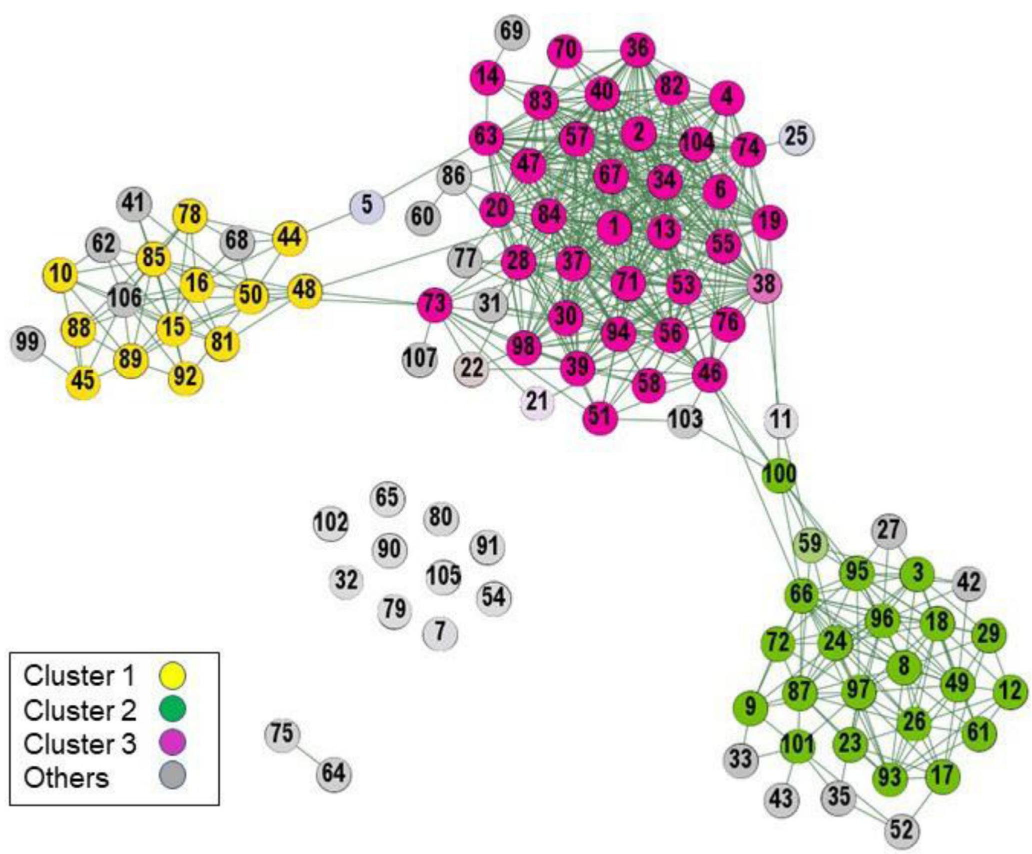

Fig. 5. The clustering structure of participants is represented in this graph by coloring. Each node represents a participant, and the connectivity is constructed based on commonality in the top 40 salient physiological features. Three major clusters are shown in yellow, green, and red, while the remaining nodes in gray represent individuals who do not belong to any cluster.

Table 1
The top common significant feature sets across the 3 clusters and the isolated participants. The most common feature set across a group is represented by "1"and the number increases as the feature set commonality decreases. SD and SI stands for signal-dependent and signal-independent features, respectively.

Significant features	$Corr_{SI}$	$Corr_{SD}$	$Zygo_{SI}$	$Zygo_{SD}$	$Z0_{SI}$	EDA_{SI}	BP_{SI}	BP_{SD}	ECG_{SD}
C1 (15 participants)	1	4	–	–	2	3	–	5	–
C2 (22 participants)	–	–	–	4	1	2	3	6	5
C3 (37 participants)	–	–	1	4	2	3	–	5	–
Others	–	3	–	4	1	2	–	5	–

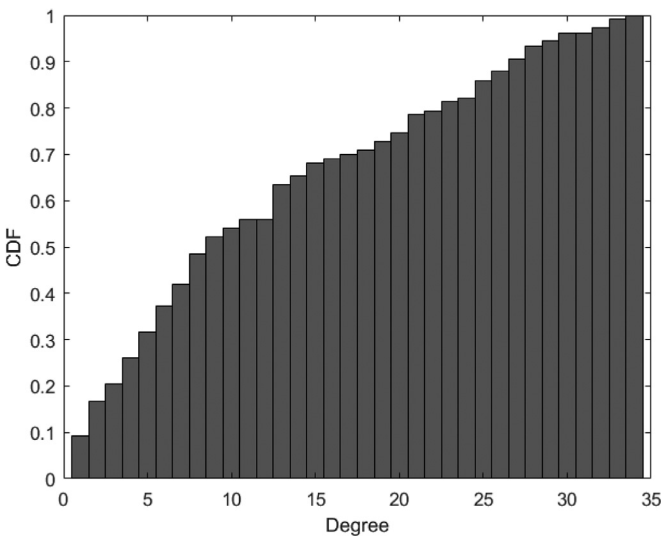

Fig. 6. Cumulative distribution function for the degree of the graph constructed using 40 features.

consistent with the existing literature that posits features of cardiovascular reactivity as most important for distinguishing biopsychosocial states during motivated performance.

3.3. *Testing Hypothesis 3: a generalization analysis*

As described in Section 2.3.2 and based on the subgroups/clusters obtained in Section 3.2, a classification/prediction analysis was performed using the self-report ratings obtained at the end of Task 3 to test our hypothesis that subjective experiences of threat and challenge would be better predicted within each subgroup than across the entire group of participants. The results are presented in Table 2. In this table, for each cluster the prediction accuracy and prediction sensitivities results are presented for 4 different inference models, $\mathfrak{T}_i$, $\mathfrak{T}_{ii}$, $\mathfrak{T}_{iii}$, and $\mathfrak{T}_{iv}$. Here, (a) $\mathfrak{T}_i$ is the model used to obtain within-cluster prediction results; (b) $\mathfrak{T}_{ii}$ is the model used to analyze the generalization of the prediction accuracy from one cluster to outside that cluster; and (c) the models $\mathfrak{T}_{iii}$ and $\mathfrak{T}_{iv}$ are trained to investigate the generalization of the prediction accuracy from outside one cluster to inside that cluster. Moreover, in Table 2, the model $\mathfrak{T}_v$ evaluates the self-report rating prediction accuracy over the entire group without considering subgroup characteristics.

Table 2
The accuracy of different inference models $\mathfrak{T}$'s to predict self-report rating SR for 3 clusters.

Inference models	Sensitivity ($SR \le 1$)	Sensitivity ($SR > 1$)	Prediction accuracy
$C1{:}\mathfrak{T}_i$	76.33%	74.47%	75.46%
$C1{:}\mathfrak{T}_{ii}$	61.70%	35.70%	50.24%
$C1{:}\mathfrak{T}_{iii}$	87.50%	0.00%	46.67%
$C1{:}\mathfrak{T}_{iv}$	80.05%	0.30%	42.84%
$C2{:}\mathfrak{T}_i$	88.01%	71.18%	79.59%
$C2{:}\ \mathfrak{T}_{ii}$	64.52%	28.52%	49.04%
$C2{:}\mathfrak{T}_{iii}$	80.27%	14.31%	47.29%
$C2{:}\mathfrak{T}_{iv}$	91.48%	08.89%	50.19%
$C3{:}\mathfrak{T}_i$	69.15%	49.06%	60.00%
$C3{:}\mathfrak{T}_{ii}$	49.15%	43.99%	46.90%
$C3{:}\mathfrak{T}_{iii}$	67.45%	43.00%	56.30%
$C3{:}\mathfrak{T}_{iv}$	100.00%	02.88%	55.38%
$All{:}\mathfrak{T}_v$	84.79%	14.45%	53.53%

In Table 2, we observe that the best self-report rating prediction accuracy was achieved when the prediction is performed within each specific subgroup/cluster with participants sharing common salient features (inference model $\mathfrak{T}_i$). It can be also observed that when the inference model is trained within a subgroup and applied to participants outside that subgroup (inference model $\mathfrak{T}_{ii}$), the prediction accuracy of that model outside the subgroup compared to the within-subgroup accuracy dropped significantly. Moreover, for a specific subgroup sharing common features, when an inference model is trained using these shared features obtained from the participants outside this subgroup and applied to the participants inside the subgroup (inference models $\mathfrak{T}_{iii}$ and $\mathfrak{T}_{iv}$), the generalization of this inference model from outside to within the subgroup was very poor compared to within subgroup prediction. Finally, when an inference model is trained within the entire sample of participants using the union of features shared by this sample and applied to the entire sample (inference model $\mathfrak{T}_v$), the self-report rating prediction accuracy is lower than the within subgroup prediction accuracy. In fact, the accuracy of $\mathfrak{T}_v$) is barely above chance (at 53.53%). Please also note that within-cluster inference models clearly out-performed all other models in terms of their sensitivity for predicting subjective threat experiences (i.e., $SR > 1$), while sensitivity for predicting subjective challenge experiences (i.e., $SR \le 1$) was consistently high across all inference models (or, in some cases, lower for within-cluster inference models). One possible reason for this outcome would be that the physiological features that best predict subjective challenge experiences are similar across all participants, while the physiological features that best predict subjective threat experiences differ across participants (and thus drive the clustering results).

4. Discussion

Classical (signal-specific) features of autonomic nervous system responding, particularly cardiovascular changes, have commonly been used to identify threat and challenge during motivated performance (e.g., Tomaka et al., 1993, 1997; Mendes et al., 2008; Jamieson et al., 2012; Quigley et al., 2002). In this study, it was found that measures of facial muscle activity (fEMG signals) and signal-independent features of the various peripheral physiological measures identified these biopsychosocial states during motivated performance. To compare the present approach to one more common approach in the existing psychophysiological literature on motivated performance, we conducted a similar set of analyses excluding both fEMG signals (COR and ZYG) as well as all signal-independent features from the remaining physiological measures. That is, we conducted the analyses using only the commonlyutilized signal-dependent features derived from measures of autonomic nervous system activity.

Using these 'classical' features, a maximum average accuracy of 88.23% was obtained when performing the within-individual analysis (classification of within-person patterns of physiological responding across the tasks of varying difficulty), which represents a reduction of approximately 9% in accuracy compared to our primary analyses with a more inclusive set of features. Moreover, the common salient features for differentiating physiological responding across tasks were the same for all participants (i.e., all participants formed a single cluster) when testing H2 with only the 'classical features'. Finally, and most critically, prediction accuracy for post-task self-report ratings of threat and challenge based on the union of salient physiological features for the full sample was as low as 48%, and alternative inference models were not able to be tested using this approach since all participants formed a single cluster when using only 'classical' physiological features. The results of this alternative analysis highlight the advantages of the analytic approach taken in the present paper: including additional measures of peripheral physiological activity (e.g., fEMG) and signal-independent features of physiological signals enables the identification of heretofore unidentified individual differences among subjects in terms of which features of their physiological activity differ most across tasks of varying difficulty. Moreover, by taking into account these newly identified individual differences, we are able to build inference models with substantially higher predictive power for post-task self-reports of threat and challenge experience compared to inference models built ignoring these individual differences (See Table 2).

Moreover, the individual differences revealed here lay bare a number of exciting avenues for future inquiry, particularly concerning the interpretation and generalizability of the observed patterns of physiological activity. For more than half of the participants in the present study (approximately 2/3), it was found that the peripheral physiological features that best differentiated within-person responding across a motivated performance task of varying difficulty were derived from measures of facial muscle activity. This is a novel discovery when compared to the existing literature and theory on understanding physiological responding in motivated performance contexts. Using these fEMG-derived features significantly improved the prediction accuracy of inference models for post-task self-report ratings of threat and challenge experience. One possibility to explore in future work is whether facial muscle activity co-varies with performance on the mental arithmetic tasks used here, such that people move their faces less as the task difficulty increases and/or as they experience more threat relative to challenge because they are attempting less subtractions per minute (i.e., they are speaking less). The present work also maximized the capacity to predict self-reports of the subjective experiences of threat and challenge collected after the tasks were completed. Much of the existing research has focused on apriori (i.e., pre-task) self-reports of threat and challenge as driving subsequent physiological activity. Thus, another interesting question to explore in future work is whether facial muscle movements are more relevant to post-task ratings of subjective experience than pre-task threat and challenge experiences, while the opposite may be true of cardiovascular variables. This prediction is consistent with previous work suggesting facial muscle activity, particularly in the corrugator supercilii and zygomaticus major muscle groups measured here, may be more predictive of the subjective experience of affective valence (i.e., feelings of pleasantness and unpleasantness) than are autonomic measures like cardiovascular activity or EDA.

Future work should also examine whether the specific salient features identified here generalize to other active coping stressor tasks and experimental contexts. For example, to the extent that fEMG-derived features are particularly relevant in the mental arithmetic task only because it involves speaking as part of task performance, it is possible that changes in fEMG-derived features would not be as relevant in motivated performance contexts that did not involve speaking (e.g., test taking, athletic performance). Moreover, previous research has demonstrated that there are significant patterns of cardiovascular adaptation to repeated active coping tasks (e.g., to completing multiple mental arithmetic tasks) (Kelsey, Blascovich, Tomaka, Leitten, Schneider, & Wiens, 1999; Kelsey, Blascovich, Leitten, Schneider, Tomaka & Wiens, 2000; Kelsey, Soderlund, & Arthur, 2004; Kelsey, Ornduff, & Alpert, 2007). As such, cardiovascular adaptation over the series of mental arithmetic tasks may contribute to the overall pattern of results. Similarly, in the present study, participants all completed numerous additional tasks prior to completing the mental arithmetic task, including an acoustic startle task and an evocative picture and sound rating task, which may have themselves been perceived as stressful to some participants. As such, prior exposure to these additional potentially stressful tasks may have led to a pattern of cardiovascular adaptation that explains why commonly-studied cardiovascular measures (e.g., CO, TPR) were not the most salient physiological features in the present investigation of the motivated performance task. Moreover, individual differences in perceptions of the stressfulness of these prior tasks may be a critical driver of the observed individual differences in salient physiological features during the mental arithmetic task. Future work should examine these possibilities.

5. Conclusion

The present study examined multimodal physiological response patterns across motivated performance tasks of varying difficulty, with the main focus on variation within and across individuals. We designed a motivated performance task that varied in difficulty level, collected the multimodal peripheral physiological activity of participants during these tasks, and performed an intensive machine learning analysis. Results revealed that there are salient physiological features which dominate the differentiation of response patterns within-individuals across mental arithmetic tasks at varying levels of difficulty. Moreover, individual differences were observed in terms of which physiological features were most salient for differentiating responding across tasks, and we identified three groups of individuals, each of which shared common salient physiological features. Experimental results using machine learning algorithms showed that the classification accuracy for predicting post-task self-report ratings of challenge v. threat experiences (i.e., the ratio of self-report stress and coping experiences) was dramatically improved by considering an individually-tailored set of physiological features compared to using an identical set of features for all participants. Future work should enhance the understanding of temporal dynamics of physiological responding during challenge and threat using multimodal data fusion and temporal network approaches.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Experimental Procedure: After reading and signing an informed consent document, the participant filled out brief questionnaires documenting health and demographic information to ensure eligibility. An experimenter measured the participant's height and weight and the participant was then instrumented for physiological recordings as noted above. Following electrode placement, the participant completed several tasks unrelated to the current investigation, including an acoustic startle reactivity task, an evocative picture and sound rating task, and a heartbeat detection task. S/he was seated in a sound-attenuated testing room in an upholstered chair throughout the experimental session. Immediately before completing the mental math task pertinent to the current investigation, the participant was connected to the continuous blood pressure monitor, and the continuous measurement from the finger cuffs on the participant's left hand was calibrated against a single non-invasive reading taken from a cuff on the right arm. The participant then sat quietly alone in the testing room while resting physiological measures were recorded for 2 min. In order to ensure this resting baseline was not confounded with anticipatory stress related to the upcoming mental math task, the participant was not informed that s/he would be doing a math task immediately after the baseline. The participant then completed the mental math task as described in Section 2.2.2. Following this, all electrodes were removed, the participant completed a set of questionnaires unrelated to the present investigation, and was then debriefed and compensated for participating.

Credit authorship contribution statement

Aya Khalaf: Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Mohsen Nabian:** Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Miaolin Fan:** Data curation, Formal analysis, Methodology, Software, Writing - original draft. **Yu Yin:** Data curation, Formal analysis, Software, Visualization, Writing - original draft. **Jolie Wormwood:** Conceptualization, Data curation, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Karen S. Quigley:** Conceptualization, Data curation, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Lisa Feldman Barrett:** Conceptualization, Investigation, Methodology, Resources, Supervision, Writing - review & editing. **Murat Akcakaya:** Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Writing - original draft, Writing - review & editing. **Chun-An Chou:** Conceptualization, Investigation, Methodology, Supervision. **Sarah Ostadabbas:** Conceptualization, Data curation, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing.

References

Sherwood, A., Allen, M. T., Fahrenberg, J., Kelsey, R. M., Lovallo, W. R., & Van Doornen, L. J. P. (1990). Methodological guidelines for impedance cardiography. *Psychophysiology, 27*(1), 1–23.

Benedek, J., & Hazlett, R. (2005). Incorporating facial EMG emotion measures as feedback in the software design process. *Proceeding of the Human Computer Interaction Consortium*.

Behnke, M., & Kaczmarek, L. D. (2018). Successful performance and cardiovascular markers of challenge and threat: A meta-analysis. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 130*, 73–79.

Berry Mendes, W., Gray, H. M., Mendoza-Denton, R., Major, B., & Epel, E. S. (2007). Why egalitarianism might be good for your health: Physiological thriving during stressful intergroup encounters. *Psychological Science, 18*(11), 991–998.

Biosignal-Specific Processing (Bio-SP) Tool. (2018). (Accessed: 2018) http://www.northeastern.edu/ostadabbas/software/.

Blascovich, J., & Mendes, W. (2000). *Challenge and threat appraisals: The role of affective cues. inj. forgas (ed.), feeling and thinking: The role of affect in social cognition (pp. 59–82)*. Cambridge, UK: Cambridge University Press.

Blascovich, J., Mendes, W. B., Hunter, S. B., Lickel, B., & Kowai-Bell, N. (2001). Perceiver threat in social interactions with stigmatized others. *Journal of Personality and Social Psychology, 80*(2), 253.

Blascovich, J., & Tomaka, J. (1996). The biopsychosocial model of arousal regulation. *Advances in Experimental Social Psychology, 28*, 1–51.

Bozhkov, L., Georgieva, P., Santos, I., Pereira, A., & Silva, C. (2015). EEG-based subject independent affective computing models. *Procedia Computer Science, 53*, 375–382.

Bsoul, M., Minn, H., Nourani, M., Gupta, G., & Tamil, L. (2010). Real-time sleep quality assessment using single-lead ECG and multi-stage SVM classifier. In *Engineering in medicine and biology society (EMBC), 2010 annual international conference of the IEEE* (pp. 1178–1181).

Chavan, M. S., Agarwala, R. A., & Uplane, M. D. (2005). Digital elliptic filter application for noise reduction in ECG signal. In *4th WSEAS international conference on electronics, control and signal processing* (pp. 58–63).

Chen, J., Hu, B., Wang, Y., Dai, Y., Yao, Y., & Zhao, S. (2016). A three-stage decision framework for multi-subject emotion recognition using physiological signals. In *Bioinformatics and biomedicine (BIBM), 2016 ieee international conference on* (pp. 470–474).

Chen, L. S., Tao, H., Huang, T. S., Miyasato, T., & Nakatsu, R. (1998). Emotion recognition from audiovisual information. In *Multimedia signal processing, 1998 IEEE second workshop on* (pp. 83–88).

De Chazal, P., Heneghan, C., Sheridan, E., Reilly, R., Nolan, P., & O'Malley, M. (2003). Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. *IEEE Transactions on Biomedical Engineering, 50*(6), 686–696.

De Luca, C., Gilmore, L., Kuznetsov, M., & Roy, S. (2010). Filtering the surface emg signal: Movement artifact and baseline noise contamination. *Journal of Biomechanics, 43*(8), 1573–1579.

Dienstbier, R. A. (1989). Arousal and physiological toughness: implications for mental and physical health. *Psychological review, 96*(1), 84.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*(4), 384.

Figner, & Murphy (2011). Using skin conductance in judgment and decision making research. *A handbook of process tracing methods for decision research*. New York, NY: Psychology Press.

Goertz, A. (1995). Left ventricular ejection time and end-systolic pressure revisited. *Anesthesia & Analgesia, 81*(4), 890–891.

Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytologist, 11*(2), 37–50.

Jamieson, J. P., Mendes, W. B., Blackstock, E., & Schmader, T. (2010). Turning the knots in your stomach into bows: Reappraising arousal improves performance on the gre. *Journal of Experimental Social Psychology, 46*(1), 208–212.

Jamieson, J. P., Nock, M. K., & Mendes, W. B. (2012). Mind over matter: Reappraising arousal improves cardiovascular and cognitive responses to stress. *Journal of Experimental Psychology: General, 141*(3), 417.

Jamieson, J. P., Peters, B. J., Greenwood, E. J., & Altose, A. J. (2016). Reappraising stress arousal improves performance and reduces evaluation anxiety in classroom exam situations. *Social Psychological and Personality Science, 7*(6), 579–587.

Kelsey, R. M., Blascovich, J., Leitten, C. L., Schneider, T. R., Tomaka, J., & Wiens, S. (2000). Cardiovascular reactivity and adaptation to recurrent psychological stress: The moderating effects of evaluative observation. *Psychophysiology, 37*(6), 748–756.

Kelsey, R. M., Blascovich, J., Tomaka, J., Leitten, C. L., Schneider, T. R., & Wiens, S. (1999). Cardiovascular reactivity and adaptation to recurrent psychological stress: Effects of prior task exposure. *Psychophysiology, 36*(6), 818–831.

Kelsey, R. M., & Guethlein, W. (1990). An evaluation of the ensemble averaged impedance cardiogram. *Psychophysiology, 27*(1), 24–33.

Kelsey, R. M., Ornduff, S. R., & Alpert, B. S. (2007). Reliability of cardiovascular reactivity to stress: Internal consistency. *Psychophysiology, 44*(2), 216–225.

Kelsey, R. M., Soderlund, K., & Arthur, C. M. (2004). Cardiovascular reactivity and adaptation to recurrent psychological stress: Replication and extension. *Psychophysiology, 41*(6), 924–934.

Kim, J., & André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(12), 2067–2083.

Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short--term monitoring of physiological signals. *Medical and Biological Engineering and Computing, 42*(3), 419–427.

Knowledge Base Impedance Cardiography. (n.d.). (Accessed: 2017) https://support.mindwaretech.com/article-categories/imp/.

Maceira, A., Prasad, S., Khan, M., & Pennell, D. (2006). Normalized left ventricular systolic and diastolic function by steady state free precession cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance, 8*(3), 417–426.

Mendes, W. B., Blascovich, J., Hunter, S. B., Lickel, B., & Jost, J. T. (2007). Threatened by the unexpected: physiological responses during social interactions with expectancy-violating partners. *Journal of Personality and Social Psychology, 92*(4), 698.

Mendes, W. B., Blascovich, J., Lickel, B., & Hunter, S. (2002). Challenge and threat during social interactions with white and black men. *Personality and Social Psychology Bulletin, 28*(7), 939–952.

Mendes, W. B., Blascovich, J., Major, B., & Seery, M. (2001). Challenge and threat responses during downward and upward social comparisons. *European Journal of Social Psychology, 31*(5), 477–497.

Mendes, W. B., Major, B., McCoy, S., & Blascovich, J. (2008). How attributional ambiguity shapes physiological and emotional responses to social rejection and acceptance. *Journal of Personality and Social Psychology, 94*(2), 278.

Pan, J., & Tompkins, W. J. (1985). A real-time qrs detection algorithm. *IEEE Transactions on Biomedical Engineering*, (3), 230–236.

Perez-Rosero, M.S., Rezaei, B., Akcakaya, M., & Ostadabbas, S. (2016). Decoding emotional experience through physiological signal processing. arXiv:1606.00370 [cs.HC].

Qu, M., Zhang, Y., Webster, J. G., & Tompkins, W. J. (1986). Motion artifact from spot and band electrodes during impedance cardiography. *IEEE Transactions on Biomedical Engineering, 33*(11), 1029–1036.

Quigley, K. S., Barrett, L. F., & Weinstein, S. (2002). Cardiovascular patterns associated with threat and challenge appraisals: A within-subjects analysis. *Psychophysiology, 39*(3), 292–302.

Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics, 23*(19), 2507–2517.

Seery, M. D. (2013). The biopsychosocial model of challenge and threat: Using the heart to measure the mind. *Social and Personality Psychology Compass, 7*, 637–653.

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., et al. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological Bulletin, 144*(4), 343.

Tkach, D., Huang, H., & Kuiken, T. A. (2010). Study of stability of time-domain features for electromyographic pattern recognition. *Journal of Neuroengineering and Rehabilitation, 7*(1), 21.

Tomaka, J., Blascovich, J., Kelsey, R. M., & Leitten, C. L. (1993). Subjective, physiological, and behavioral effects of threat and challenge appraisal. *Journal of Personality and Social Psychology, 65*(2), 248.

Tomaka, J., Blascovich, J., Kibler, J., & Ernst, J. M. (1997). Cognitive and physiological antecedents of threat and challenge appraisal. *Journal of Personality and Social Psychology, 73*(1), 63.

Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics, 15*(2), 70–73.

Wu, D., & Parsons, T. D. (2011). Inductive transfer learning for handling individual differences in affective computing. In *Affective computing and intelligent interaction* (pp. 142–151). Springer.

Young, S. S. (2001). *Computerized data acquisition and analysis for the life sciences: A hands-on guide*. Cambridge University press.