A Multimodal Dataset and Evaluation for Feature Estimators of Temporal Phases of Anxiety

Hashini Senaratne
Levin Kuhlmann
Kirsten Ellis
Faculty of IT, Monash University
Australia

Glenn Melvin
School of Psychology,
Deakin University
Australia

Sharon Oviatt
Faculty of Engineering,
Monash University
Australia

ABSTRACT

Vicious cycles of anxiety responses underlie the onset of increasingly prevalent and highly impairing anxiety disorders and also contribute to their maintenance. Our goal is to evaluate whether different anxiety responses are evident in temporal patterns of physiological and behavioral features. Consequently, we established a rich multimodal-multisensor dataset of cardiac, electrodermal, movement, posture, and speech measures from 95 young adults during two anxiety experiments that induce social anxiety and bug-phobic anxiety. A subset of this dataset is publicly available at "Anxiety Phases Dataset" Figshare repository. We adopted a generalized mixed model approach and found that 10 out of 14 feature trajectories modeled for high- and low-anxiety groups differ significantly at 0.001 level in magnitude, creating at least two temporal phases in both groups. Further differences in magnitude, duration and the number of phases were observed for responses of confrontation, safety behaviors, escape, and avoidance in the high-anxiety group. Our findings contribute to the long-term aim of designing multimodal systems that have great potential to reduce the impacts of anxiety disorders and improve therapy.

CCS CONCEPTS

• Human-centered computing; **• Applied computing → Health informatics**; **• Computing methodologies → Modeling and simulation**;

KEYWORDS

Anxiety, Generalized additive mixed models, Multimodal-multisensor dataset, Objective detection, Temporal phases

ACM Reference Format:
Hashini Senaratne, Levin Kuhlmann, Kirsten Ellis, Glenn Melvin, and Sharon Oviatt. 2021. A Multimodal Dataset and Evaluation for Feature Estimators of Temporal Phases of Anxiety. In *Proceedings of the 2021 International Conference on Multimodal Interaction (ICMI '21), October 18–22, 2021, Montréal, QC, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3462244.3479900

ICMI '21, October 18–22, 2021, Montréal, QC, Canada

ACM ISBN 978-1-4503-8481-0/21/10…$15.00
https://doi.org/10.1145/3462244.3479900

1 INTRODUCTION

Anxiety is the irrational anticipation of uncertain future threats [3, 25]. Individuals with anxiety disorders experience excessive symptoms of anxiety over a period of time that impact upon their functioning [3]. These disorders are increasingly prevalent [71] and costly for the society [34]. Common anxiety disorders of adults include social anxiety disorder, specific phobia, panic disorder, agoraphobia, and generalized anxiety disorder, according to the Diagnostic and Statistical Manual of Mental Disorders (DSM−5) [2].

In practice, assessments of anxiety are run by clinicians at therapy sessions and are subjective in nature. These assessments are difficult to access and often delayed due to the barriers and intermittent nature of therapy [31, 34, 47]. Further, the accuracy of these assessments is affected by many biases [37], as they heavily rely on patients' subjective retrospective reports of anxiety experiences and clinicians' judgment. A growing branch of research, *objective anxiety assessment research*, aims to reduce these drawbacks by attempting to objectively detect anxiety through the analysis of physiological and behavioral measures related to anxiety [26, 71].

Objective anxiety assessment research has extensively studied the slow-changing impairment levels of anxiety disorders through the evaluations of physiological and behavioral differences between patients and healthy controls during resting periods or exposure to triggers [11, 16]. However, the constructs of rapidly-changing state anxiety experiences that contribute differently to the development and maintenance of anxiety disorders remain under-evaluated.

1.1 Theory Behind Anxiety Responses

Cognitive theories of the *ABC-model* by Albert Ellis [17] and *Generic Cognitive Model* by Aaron Beck [6] explain that anxiety is a consequence of threat appraisals. Appraisals causing anxiety are driven by underlying dysfunctional cognitive structures or schemas that take the forms of unconditional beliefs (e.g., "I am incompetent in public speaking") and conditional assumptions (e.g., "Unless I avoid spiders, they will attack me"). Individuals tend to update their appraisals as situations unfold [15]. Consequently, anxiety symptoms and their severity levels tend to fluctuate during state anxiety experiences [23]. Different responses are also likely to occur, depending on individuals' vulnerability for associated triggers determined by their anxiety traits and recent anxiety experiences [2, 18].

Revised Reinforcement Sensitivity Theory by Jeffrey Gray suggests that at least one of the three behavioral systems activates during a state anxiety experience, contributing to different responses [51]. They are behavioral activation system (*BAS*), fight-flight-freeze system (*FFFS*), and behavioral inhibition system (*BIS*). The BAS contributes to energetic and appetitive confrontation of a situation that triggers anxiety, and it activates due to individuals' motivation for reward. The FFFS contributes to fight, flight, and freeze responses, and it activates due to individuals' fear of punishment. During fight responses (i.e., a defensive confrontation), individuals often use safety behaviors, which are performed to reduce the impact of a perceived threat [15]. Flight is the defensive withdrawal (i.e., escape) from a situation after attempting to fight. Freeze responses can occur before fight or flight due to hypervigilance [4]. The BIS activates when individuals are uncertain about whether to expect a reward or a punishment. Individuals with more reactive BIS are assumed to be more anxious and often avoid perceived threats [45].

Adverse anxiety experiences create vicious cycles that fuel anxiety [19, 45], which result in repeated use of avoidance, escape and safety behaviors like responses (rather than confrontation). They reinforce the impaired learning that individuals are incapable of dealing with anxious situations, creating and maintaining dysfunctional cognitive schemas [15]. In return, they contribute to the development and maintenance of anxiety disorders. Therefore, attempts to objectively study distinct short-term anxiety responses can be related to several beneficial impacts (see Section 1.3).

1.2 Related Work

Limited studies have evaluated different types of anxiety responses or fine-grained symptom fluctuations that occur within state anxiety experiences using changes in human objective measures. Summarized below are examples of existing evaluations.

Richter, et al. have analyzed a dataset of physiological and behavioral measures from 345 patients with panic disorder and agoraphobia to evaluate avoidance, escape and confrontation responses to an agoraphobic trigger [52]. They have conducted a univariate analysis of variance on heart rate (HR), skin conductance level (SCL), and startle eyeblink magnitude for intervals of before, during, and after entering and staying at an enclosed small darkroom. HR and SCL reflect sympathetic arousal. Startle response magnitude relates to the processing level of external cues. While preparing for exposure to the trigger, HRs and SCL have been highest in avoiders and lowest in confronters who perceived low state anxiety levels. During the exposure interval, escapers have experienced the highest HRs and startle response magnitudes. Confronters who perceived high state anxiety levels have experienced the highest SCL during exposure and recovery intervals. This study has also provided minute-to-minute mean feature trends for some responses, indicating high sympathetic arousal and inhibited processing of external cues immediately before exposure in escapers. While it has not provided an evaluation of the differences between responses for each time point, overall results indicate that evaluated measures fluctuate during anxiety experiences differently for different anxiety responses.

Rosenfield, et al. have evaluated 13 naturally occurring panic attacks of 11 subjects using change-point analyses on mean time-series of HR, SCL, physical activity levels and speech occurrence [44, 53]. They calculated a time point at which a change point occurred using maximum likelihood estimation. They have found that repeated HR bouts occur from ~30 minutes prior to a panic attack, and the onset of a panic attack is marked by a significant increase of HR with no further fluctuations during the attack. Similar cascaded change points have also occurred for several respiratory measures, where breathing signals were captured through an obtrusive sensor (a nasal cannula) as they often get affected by several external factors. Evaluated SCL measures were observed during panic attacks compared to non-panic intervals. A decrease followed by an increase in physical activity has also been identified before panic attacks. No change points were discovered in speech occurrence. Overall, this study also confirms that physiological and behavioral features produce temporal patterns within anxiety experiences.

In accordance with the findings of the above studies, a recent interview study conducted with anxiety specialist clinicians has suggested that four temporal phases likely occur within diverse types of anxiety experiences [56]. Suggested phases are non-anxious, pre-anxious, peak-anxious, and post-anxious phases. These four phases were expected to be cyclic across anxiety experiences with potential variations in order and duration depending on factors like trigger types and other individual differences in anxiety (e.g., responses of confrontation, safety behaviors, escape, and avoidance). The clinician participants have suggested several suitable treatment components to be used within postulated phases (e.g., relaxation or cognitive restructuring exercises during anticipatory intervals), indicating that in-the-moment digital interventions can be developed if the real-time detection of such phases is possible. Although they have hypothesized that such phases would be detectable through multimodal feature analyses, such evaluations are yet to emerge.

While existing evaluations have focused mainly on anxiety experiences of patients with panic disorder (with or without agoraphobia) [44, 52, 53], such samples often exhibit highly contrasting symptoms rather than samples with other anxiety presentations. Similar evaluations are required for anxiety experiences involving social anxiety, specific phobia and generalized anxiety triggers. Further, evaluation of multimodal estimators of temporal phases of anxiety require comparisons between (1) patient and healthy subjects to explore whether they experience different phases, (2) different triggers to explore whether they contribute to the occurrence of varying phases, and (3) diverse features to postulate which may carry complementary information about occurring phases.

1.3 Hypotheses, Novelty and Contributions

The aim of this research is to evaluate and compare temporal patterns of anxiety, which are evident in a variety of physiological and behavioral feature changes and occur within two distinct anxiety experiences. We evaluate two main hypotheses:

- `HP1`: Temporal patterns of anxiety occur across relaxing, anticipatory, exposure and recovery intervals, with differences between high-anxiety and low anxiety groups.
- `HP2`: Temporal patterns in high-anxious individuals vary depending on anxiety responses, independent of trigger types.

A significant novel contribution of the present research is the established multimodal dataset. It includes subjects representing high- and low- anxiety groups, who went through a pair of common anxiety experiences within a controlled setting with relation to public speaking and performance anxiety (a type of social anxiety), bug phobic anxiety (a type of specific phobia), and generalized anxiety. It also covers a range of anxiety responses and readings of cardiac, electrodermal, movement, posture and speech data.

Another contribution is the utilized short-term evaluation method. It is cost-effective and could be used to identify anxiety-prone candidates for longer-term monitoring or psychological care.

Further, our findings propose the ability to objectively detect distinct state anxiety experiences, contributing towards the long-term aims of minimizing the burden of anxiety disorders. For example, real-time detection capabilities can support implementing intelligent mobile technologies to guide appropriate and timely interventions to patients outside therapy [8], ultimately supporting to break the vicious cycles of anxiety, therefore; improving the recovery speed or preventing anxiety disorders [56]. Longitudinal analytics on temporal pattern changes across anxiety experiences could estimate risk levels and guide the general public to clinicians on time to receive early support. Clinicians could use such estimates to assess patients' progress and tailor treatments accurately. Progress in this direction also has the potential to shed new insights on how anxiety disorders develop and progress over time.

2 DATA COLLECTION

2.1 Participant Pool

Ninety-five young adults aged 18-35 (52.63% female) were recruited through advertising at a university, psychology clinics and local social media groups, including mental health support groups. Participants were South Asian (47.37%), East Asian (27.37%), Caucasian (23.16%), and Black or African American (2.1%). All were paid volunteers. Forty-one subjects (63.41% females) had recently sought professional or specialized support for anxiety issues. Others (44.44% females) were free of anxiety disorders and had no history of receiving external support or medication for anxiety. All were free of other clinically diagnosed mental health disorders; however, comorbid depression was common in the group who received support.

To reduce unknown effects on physiological and behavioral measures, we excluded participants based on a number of criteria. Those are: usage of medication such as beta-blockers, substance dependence, motor disabilities, clinically diagnosed cardiovascular and neurological diseases, non-English speaking (as it would impede them from following instructions), vision or hearing impairments without corrections and intellectual disabilities. We also excluded people with diagnosed panic disorders as a risk mitigation step. Exclusions were based on a self-reported pre-study questionnaire.

2.2 Study Procedure

The session duration was closer to one hour (see Figure 1). The first author conducted studies with institutional ethics approval.

2.2.1 Before the Study: Participants were requested to schedule a timeslot so that they could refrain from consuming drugs, medication or stimulants on the study day, aiming to minimize unknown effects on measures. There was no mention of study task content, so it remained novel to participants. However, they were informed that it was an emotional awareness study that uses wearables.

2.2.2 On Arrival to the Study: Participants were seated in front of a computer system with a monitor, keyboard, wireless mouse, and headphones. The researcher first introduced an HTML-based interface (with audio-visual instructions and questions) that was developed to guide participants and collect their inputs. It first collected their consent and guided them on how to wear sensors.

After ensuring that wearables were worn correctly, the researcher left the study room, allowing participants to proceed with system instructions. Participants had the option to call the researcher through the system asking for assistance. The researcher also observed participants remotely without their knowledge. A psychologist was also on standby to provide any assistance in managing distress.

2.2.3 Baseline Intervals: The system instructed participants to stay at rest for two minutes before moving to the tasks, and sensor data collected during this interval are considered as baseline data for each metric, except speech. Next, they were asked to verbally describe what they can see in a neutral picture shown on the screen for two minutes. The picture was a living room with a number of objects and without people. This interval collected speech baselines.

2.2.4 Study Tasks: The study consisted of two within-subject tasks, which were counterbalanced and designed to induce two prevalent anxiety presentations in young adults [32]: bug phobia [64] and public speaking or performance anxiety [69]. Each task provided five time-intervals to: (1) calm down by watching and listening to a nature scene with soothing music, (2) prepare to engage with a trigger, (3) expose to the trigger, (4) recover, and (5) reflect the experience in speech. Each interval was approximately 3 minutes, a commonly used interval in similar study protocols [42]. Each task had some level of uncertainty incorporated and space for decision making, aiming to trigger generalized anxiety as well [7, 25].

During the *bug phobic task*, after the calm interval, the system provided instructions to find a bug-box that was covered by a cardboard box on the table and asked to pick it up and prepare to release the bug inside. The bug-box was a wooden box with laser-cut images of bugs, air vents, and warnings (see Figure 2a). It had circuitry inside to generate a random vibration when picked up. A small polythene bag was also on top of the box without any instructions for its use. The system also did not mention the kind of bug to expect. After preparation, participants were asked to release the bug by directly opening the door or while covering the door with the polythene bag. The circuitry inside the box was designed to move a toy cockroach slightly forward when the door was opened.

During the *speech anxiety task*, after the calm interval, the system provided three topics ("Climate change is a natural occurrence", "My

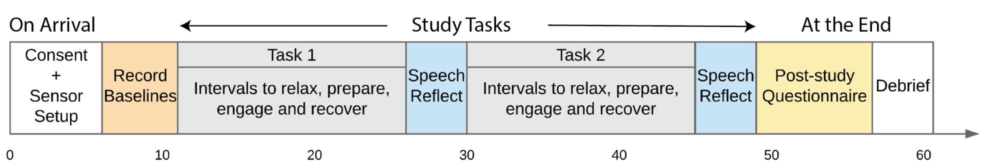

Figure 1: Study Timeline.

biggest concern for the future is...", and "Facebook makes society less happy") and required sorting them in the order of difficulty. Then, participants were asked to prepare for a 3-minute speech on the most difficult topic, providing space on the screen to take down notes. The system did not mention the type of audience. After preparation, they were asked to deliver the speech to a video of a large audience (see Figure 2b) while viewing the notes or not. The system further mentioned that the research team is going to watch and evaluate the speech recordings at a later stage.

Except for the trigger differences, both tasks had similar components. The system displayed a countdown of 3 minutes during preparation intervals. Participants could further request an additional minute to prepare. After preparation, they also had the option to skip or stop proceeding with the main goal (i.e., releasing the bug or delivering the speech) at any time they wanted. Then, they were asked to spend three minutes freely. Finally, they were requested to describe how they were feeling and what they were doing during each task. This interval is not included in the presented analysis.

2.2.5 At the end of the Study: Participants were requested to fill a post-study questionnaire (see Section 2.3.2 for details). Finally, the researcher came in and debriefed on the full nature of the study.

2.3 Collected Data

2.3.1 Multimodal-multisensor data. During the study, an array of sensors recorded physiological and behavioral data of participants.

A Zephyr BioHarness 3.0 bio patch [72] was worn using two disposable adhesive electrodes aligned to the breast bone. It collected electrocardiogram (*ECG*) data at a sampling rate of 250Hz. It also recorded activity levels (in VMU) and posture (in degrees) at 1Hz.

A Grove-GSR Sensor V1.2 [55] and a WitMotion WT901BLECL BLE 5.0 [68] enclosed in a 3D-printed module with circuitry to record data into an SD card was worn on both wrists using Velcro straps. Two electrodes of each GSR sensor were worn at the index and ring finger. These sensors collected skin resistance derivable data and 9-axis motion data (i.e., triaxial acceleration, angular velocity and inclination angles) at a sampling rate of 50Hz. Similar modules without GSR sensors were worn on the two ankles.

Audio was digitally recorded using a lapel microphone clipped to the collar area of participants and a Philips DPM8900 recorder.

A webcam captured the face and upper torso areas, and a 360-degree camera attached to a 2m tripod captured an overall view. The researcher used the second video stream for remote observation.

Figure 3 illustrates sensor positioning within study setup. All sensors recorded each data entry with a timestamp, except for the audio and video recorders. The researcher used a digital annotator with no more than 1-second error to stamp their starting times.

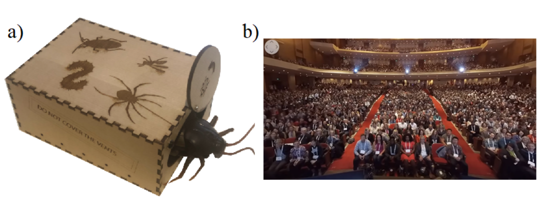

Figure 2: Triggers Used: a) Bug-box, b) Large Audience Video.

2.3.2 Subjective Anxiety Scores. Within 2 days prior to the study, participants filled an online version of the 42-item self-report Depression Anxiety Stress Scale (DASS) [49]. It estimates the severity of depression, anxiety and stress of the past week, with considerably high internal consistency and discriminant validity [1]. It captures 5 general anxiety severity ranges: normal: 0-7, mild: 8-9, moderate: 10-14, severe: 15-19 and extremely severe: 20+.

During the post-study questionnaire, participants rated state anxiety levels experienced during calm, anticipate, expose and recover intervals within each task. For this, 10-item (range: 0-100) Subjective Units of Distress scale [43] was used. They also rated 5 items of the Leibowitz Social Anxiety Scale (LSAS) [28] on public speaking anxiety and avoidance, and 5 equivalent items formulated for bug phobic anxiety and avoidance. Scores of anxiety sub-scales (the avoidance subscale is not used in this analysis) provided estimates for participants' trait anxiety severity for each anxiety type.

3 DATA PRE-PROSSESSING

Pre-processing was done in Python. As a synchronization step, all sensor data streams were formatted to "epoch timestamp: value".

3.1 Data Cleaning

ECG signals were first enhanced [63], then filtered with a high-pass filter (cut-off = 0.667Hz, order = 2) and an 8-point moving average filter to reduce noise that can be related to breathing and muscle activity [33]. Finally, two bandpass filters were applied (cut-off-1 = 8-50 Hz, cut-off-2 = 0.75-2.5 Hz, order = 3) to isolate RR intervals of ECG signals. Their cut-offs matched with the frequency range of QRS peaks of ECG signals and valid range of RR intervals [61].

The Grove sensor readings from the non-dominant hand, which had lesser motion artifacts, were chosen over the dominant hand when available. Readings were first passed through a low-pass filter (cut-off = 1Hz, order = 3) [50]. Filtered readings (x) were then converted to skin resistance as SR = 200*(272+x)/(752-x) kΩ; where this equation was achieved by solving the underlying circuitry and verified through simulation. Skin conductance is derived as SC = 1/SR. It was then decomposed into tonic component of Skin Conductance Level (SCL) by passing data through a median value smoothing filter (smoothing factor = 4) [41]. Phasic component of Skin Conductance Responses was then derived as SCR = SC - SCL.

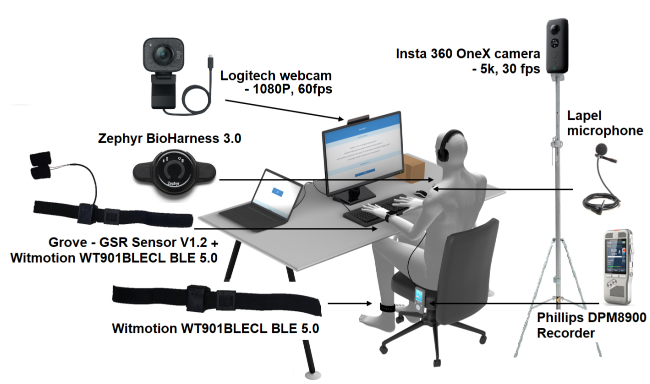

Figure 3: Sensors Used and Study Setup

Posture and activity data of the torso were directly extracted from the Zephry sensors. Other triaxial acceleration data of wrists and ankles were passed through Butterworth high pass filters (cut-off = 0.8) to isolate the gravity effect. As these sensors had an integrated Kalman Filtering technique, no other filtering was applied.

Each audio file was filtered for continuous noise using Fast Fourier Transformation (FFT) based masking method suggested in the "noisereduce" python package [54]. A 60-second segment from the baseline interval was used as sample background noise.

3.2 Data Discards

Through visual inspections, 8 ECG, 13 SC and 2 ankle motion streams (from 18 participants) were identified as noisy or invalid due to loosely connected or malfunctioned sensors. Further, 8 bug-box and 2 speech tasks had instances of calling for the researcher's assistance or a malfunctioned system or bug-box. These 18 subjects and 10 task segments were discarded from subsequent steps.

3.3 Ground Truths

According to the following criteria, two annotators labeled participants' responses to each study task as avoidance, escape, safety behaviors, or confrontation based on videos at a 100% agreement. Those who did not attempt to pick up or open the bug-box or deliver the speech were considered to have avoided the task. Participants, who started opening the bug-box but did not release the bug or started delivering the speech but stopped at least 20 seconds before the finish time, were considered to have escaped the task. Safety behaviors were operationalized as using an extra minute to prepare, using the polythene bag to cover the door while opening the bug-box or using speech notes. Those who completed the task without demonstrating the above responses were tagged as confronters.

To form high-anxiety (HA) and low-anxiety (LA) groups, we used a convergent ground truth, aiming to reduce biases of subjective measures. For example, ones who do not seek external support could have high anxiety impairment levels. Likewise, those who receive support could have benefited from treatments and reduced impairment levels. Therefore, those who have recently received support for anxiety and had higher DASS anxiety scores and those who had no history of receiving support and lower DASS anxiety scores were only considered to form high and low anxiety groups. This converged information was assumed to produce better ground truths in comparison to reliance on a single measure [10]. However, we did not use a static DASS anxiety cut-off for the above criteria due to dataset limitations and requiring a sufficient and balanced number of participants covering different anxiety responses.

Within each task, participants were categorized into 8 sub-groups (2 Support Levels x 4 Responses). In the bug-box task, at least 5 subjects represented each subgroup. However, in the speech task, we had only 2 subjects representing high support - escape, and 0 representing low support - avoidance, and these numbers were not sufficient to form sub-groups. Considering the highest (for forming HA group) or lowest (for forming LA group) DASS anxiety scores within these subgroups, we selected 40 subjects each for bug-box task (5 x 4 Responses x 2 Support Levels) and speech task (10 x 2 Responses of Confront and Safety Behaviors x 2 Support Levels). When we had to select between multiple subjects with similar DASS anxiety scores, the highest or lowest Leibowitz Social Anxiety Scale scores or corresponding bug phobic anxiety scores were utilized, depending on the task. Overall, 55 subjects contributed to 80 data streams extracted for two tasks. Analyzed data (except speech) is available in Figshare repository: Anxiety Phases Dataset [57]. Table 1 summarizes mean differences of subjective anxiety scores from baselines related to four intervals of each task, mean DASS anxiety scores and gender distributions of the formed subgroups.

Table 1: Sub-group wise subjective anxiety score differences from baselines, DASS anxiety scores and gender details

	Bug-Box Task				Speech Task	
High-Anxiey Sub-groups:	1	2	3	4	1	2
Mean SUDs : Calm	-20	-4	-24	-2	-7	-11
Mean SUDs : Prepare	2	22	34	40	19	22
Mean SUDs : Exposure	10	50	56	32	22	28
Mean SUDs : Recover	-12	10	20	8	0	1
Mean DASS Anxiety	9.4	16.4	9	13	11.7	13.8
% Females (100% = 40 subjects)	100	40	40	80	90	60
Low-Anxiey Sub-groups:	1	2	3	4	1	2
Mean SUDs : Calm	-20	-12	-22	-8	-25	-11
Mean SUDs : Prepare	-2	-10	8	30	2	12
Mean SUDs : Exposure	-4	4	18	28	4	16
Mean SUDs : Recover	-14	-20	-14	10	-16	-5
Mean DASS Anxiety	2.8	2.4	3.4	5.8	4	3.3
% Females (100% = 40 subjects)	20	60	60	60	20	50

Mean SUDs = Mean of Subjective Units of Distress Score Differences from Baselines.
Mean DASS Anxiety = Mean of Anxiety Scores of Depression Anxiety Stress Scale.
1: Confront, 2: Safety Behaviors, 3: Escape, 4: Avoidance

4 FEATURE EXTRACTION

We focused our analysis mainly on four intervals of two tasks (bug-box and speech). First, cleaned and selected data streams were cropped to baseline and task segments considering system recorded timestamps, where task segments consist of calm, prepare, expose and recover intervals. Then, they were also cropped of extra time spent by some participants to prepare and switch between system screens, so the final feature time-series would be of the same length.

We extracted 14 features for every 5 seconds along segmented data streams, producing feature time-series of equivalent sampling rates. Since the derived cardiac features [58], wrist rigidity and SCR rate [9] are not meaningful when calculated for ultra-short intervals, a moving-average technique was used with an overlapping window of 55 seconds. For other features extractions, no overlaps were used.

The selected features are identified as highly associated with anxiety traits, states or impairments in literature, hence assumed to be predictive of anxiety phases. Following are their descriptions, along with highly reported associations. Facial expressions are not included due to less feasibility for continuous recording in the wild, as we aim to transfer the findings to future mobile technologies.

F1: Mean of RR intervals (*Mean RR*). It inversely correlates to heart rate and sympathetic arousal of many anxiety responses [62].

F2: Short-term heart rate variability operationalized by Root Mean Squared Differences of RR Intervals (*RMSSD RR*). Reduced values reflect disturbed balance of cardiac activity [58, 62].

F3 - F4: High (0.04–0.15 Hz) and low (0.04–0.15 Hz) frequency power band of RR intervals (*LF - RR & HF - RR*) calculated by Fast Fourier Transformation. HF and LF often reflect parasympathetic and sympathetic activity prominence, respectively [58].

F5: Mean Skin Conductance Level (*Mean SCL*). It reflects slow-changing (tonic) negative or positive valenced arousal level [35].

F6: Number of Skin Conductance Responses per Minute (*SCR Rate*). It reflects short-term (phasic) arousals [35].

F7: Mean Posture, in terms of tilted angles relative to an erect posture, where positive values indicated collapsed posture. Collapsed postures are common in anxious states [38, 66].

F8: Mean activity level of the torso (*Mean Act. Torso*). Low activity has been related to high trait anxiety [39], whereas sympathetic arousal increases and freeze-like responses decreased activity levels.

F9: Mean rigidity of movements in two wrists (*Mean Rigid. Wrists*), operationalized by % of recurrent points exhibiting determinism calculated based on a Recurrent Quantification Analysis, as in [36]. Rigid movements are often related to anxiety [65].

F10: Mean activity level of ankles (*Mean Act. Ankles*), operationalized by the average of root mean squared sum of x, y and z acceleration components of two ankles. Fidgety movements commonly observed in anxiety states produce high activity levels [13].

F11-F14: *Mean Loudness* (F11); *Mean Pitch* (F12) and *Stdev Pitch* (F13) operationalized by mean and standard deviation of logarithmic F0 on a semitone frequency scale starting at 27.5 Hz; and *F14 - Speech rate* operationalized by number of continuous voiced regions per second. These were extracted as per OpenSmile eGeMAPSv01b acoustic feature set [20] only for speech delivery interval, as speech levels were minimum in other intervals. Anxiety states are often characterized by reduced loudness [5, 22], increased speech pitch [66], increased pitch variation [48], and higher speech rates [24, 59].

From each feature value, feature means calculated for corresponding baseline intervals were deducted as in [21]. This deducted value served as a human-centered baseline to isolate the individual differences in physiological and behavioral features. We also imputed empty feature values related to avoidance with last feature value of preparation interval to generate same-size feature arrays.

5 ANALYSES AND RESULTS

We use Generalized Additive Mixed Models (GAMMs), a special case of smoothing spline analysis of variance, to test the effects of anxiety severity and reactions on feature trajectory shapes and magnitudes [27, 70]. It is a statistical technique that models non-linear relationships between time-varying predictors and outcome variables. It has been used in domains such as acoustic and physiological analyses [14, 29, 60], as it combines the strengths of flexibility (as in machine-learning techniques) and interpretability (defined as an addition of linear models). GAMMs facilitate deriving a model for a cluster of time series data by applying non-parametric smoothing functions and comparing significant differences between multiple models; therefore, suitable for testing our hypotheses.

5.1 HP1: Modeled feature trajectories differ for high- and low-anxiety groups

We generated GAMMs for each feature, separately for HA and LA groups of each task, using the BAM function from the R package "mgcv", as in Equation (1). Non-linear smoothing functions (s) were applied for both time intervals and the interactions between intervals and group [60]. The basis function of the smooth terms was configured to be a cubic regression spline, and the estimation method was configured to be a maximum likelihood model.

$$Fx \sim Group + s(Interval) + s(Interval, by = Group) \quad (1)$$

Significance differences between the GAMMs of two groups for each task were tested using the *compareML* function of "mgcv", where we fed the above model and a nested model that excludes the difference smooth, as in [60]. Table 2 (1) presents the results for each feature. Except for the high- and low-frequency power band of RR intervals (F2, F4) and mean pitch (F12), all other features demonstrated 0.001 significant differences for both tasks or evaluated task intervals. Their trends are further illustrated in Figure 4.

5.2 HP2: Modeled feature trajectories differ for responses of high anxiety group

We further generated GAMMs for different responses within high-anxiety group for features that showed $p<0.001$ significant differences between HA and LA groups. For this, non-linear smoothing functions were applied for time intervals and the interactions between intervals and response, as in Equation (2). Here, we used similar configurations and a comparison method as in HP1 testing.

$$Fx \sim Response + s(Interval) + s(Interval, by = Response) \quad (2)$$

Table 2(2) presents the results of significant testing conducted for response types. For the bug-box task, all selected features demonstrated a significant difference (at 0.001) between four reactions. However, since the compareML method had no correction for multiple comparisons, we also conducted pairwise comparisons using "multcomp" R package and confirmed that at least one adverse response (avoid, escape, safety) in comparison to confrontation had a significant difference at 0.001. For the speech task, except for mean loudness (F11) and stdev pitch (F13), other features demonstrated significant differences at 0.001. See Figure 4 for feature trends.

6 DISCUSSION

Our findings confirmed that temporal patterns evident in objective features differ depending on anxiety severity and response types within social anxiety and specific phobic anxiety experiences. In fact, significant differences (at 0.001) were observed in 10/14 GAMM-based feature trajectories modeled for high- and low- groups and 8/10 GAMM-based feature trajectories modeled for response types in the high anxiety group across study tasks. Maximum and minimum patterns of many feature models also indicated that at least 2 phases occurred during preparation and exposure intervals of each task in both the groups; however, with differences in magnitude.

6.1 Magnitude Differences in Phases

The magnitude differences between GAMM models of multiple features implied that the high-anxiety group experienced higher anxiety levels than the low-anxiety group during both anxiety experiences, and comparatively higher anxiety levels were experienced in the bug-box task than the speech task, particularly during the phase that occurred in the preparation interval. These observations were closely aligned with subjective anxiety scores reported in

Table 2: Summary of significance differences between generalized additive models of each feature derived for (1) high- and low-anxiety groups for each study task and (2) response types in high-anxiety group for each study task.

Feature differences from baselines	(1) Differences of High- and Low Anxiety Groups				(2) Differences of Responses in High-Anxiety Group			
	Bug-Box Task		Speech Task		Bug-Box Task		Speech Task	
	AIC diff.	p-value Sig.	AIC diff.	p-value Sig.	AIC diff.	p-value Sig.	AIC diff.	p-value Sig.
F1: Mean RR	278.56	< 0.001***	22.93	< 0.001***	53.53	< 0.001***	9.03	< 0.001***
F2: RMSSD RR	174.52	< 0.001***	9.62	< 0.001***	174.57	< 0.001***	57.78	< 0.001***
F3: HF - RR	8.44	< 0.001***	2.13	0.235				
F4: LF - RR	11.89	< 0.001***	7.67	< 0.01**				
F5: Mean SCL	176.11	< 0.001***	343.22	< 0.001***	143.04	< 0.001***	15.64	< 0.001***
F6: SCR Rate	298.95	< 0.001***	167.85	< 0.001***	380.85	< 0.001***	33.77	< 0.001***
F7: Mean Posture	116.46	< 0.001***	45.50	< 0.001***	220.77	< 0.001***	157.74	< 0.001***
F8: Mean Act. Torso	11.34	< 0.001***	96.60	< 0.001***	31.73	< 0.001***	16.03	< 0.001***
F9: Mean Rigid. Wrists	469.70	< 0.001***	157.25	< 0.001***	15.25	< 0.001***	12.85	< 0.001***
F10: Mean Act. Ankles	3.24	0.09	55.65	< 0.001***				
F11: Mean Loudness	N/A	N/A	80.61	< 0.001***	N/A	N/A	1.726	0.33
F12: Mean Pitch	N/A	N/A	1.71	0.33				
F13: Stdev Pitch	N/A	N/A	21.89	< 0.001***	N/A	N/A	1.64	0.35
F14: Speech Rate	N/A	N/A	44.26	< 0.001***	N/A	N/A	29.44	< 0.001***

AIC diff: Difference between Akaike information criterion for compared GAMMs. Note: Last four acoustic features are only evaluated for speech exposure interval.

Table 1. Among significant features, only speech rate (F14) and standard deviation of pitch (F12) had exceptions to these interpretations, which were expected to be higher in the HA group.

Reduced heart rate variability (F2) and increased mean skin conductance levels (F5) were observed in confronters compared to ones that demonstrated safety behaviors in both tasks. These redundant observations point to modes to differentiate these two responses.

According to physiological measures of heart rate variability (F2) and skin conductance (F5-6), confronters seemed to experience the highest arousal levels most of the time, which is contradictory to perceived anxiety scores (see Table 1). Aligned to our subjective measures, a previously detailed study [52] has reported higher arousals in escapers or avoiders compared to confronters during an agoraphobic task using measures of heart rate and skin conductance. However, the current study shows that physiological changes in avoiders and escapers are more suppressed, except for limited time periods where mean RR intervals (F1) have become significant. Physiological suppression has also previously been observed in high anxious individuals during worry [30]. It is not well understood why such differences in objective measures occur for the same response in different situations. Although prediction of avoidance and escape seems challenging, decreased skin conductance response rates (F6) occurred in recovery intervals (indicating a sense of relief) can be identified as a pattern to differentiate them from other responses.

6.2 Frequency and Duration of Phases

Although many feature trajectories indicated 2 phase occurrences for both high and low anxiety groups and many reactions, some features indicated 3 phase occurrences. Instances demonstrating 3 phases are related to feature models of skin conductance response rate (F6), mean rigidity of wrists (F9) and mean RR interval (F1) for confronters of the bug-box task, and feature models of mean skin conductance level (F5) and mean torso activity level (F8) for both confronters and ones who had safety behaviors in the speech task. This additional phase is closer to the recovery period and could have occurred due to ruminations [19]. Although speech rate (F14) like acoustic features differed significantly between groups, they did not contribute to identifying phases due to their discrete nature.

The heart rate variability (F2) feature significantly increased in subjects with safety behaviors closer to the recovery period before a similar pattern occurred in confronters and escapers. This increase can be related to a sense of relief felt maybe after revealing that the bug was fake or getting closer to the end of the speech. Although difficult to depict based on Figure 4, due to imputed data, avoiders experienced a similar pattern even before, as they skipped the exposure period. Therefore, occurred phases differed in duration.

Overall, many feature trajectories of HA group differed according to their reactions with differences in magnitude, number of phases occurred and phase duration. Reasons for feature fluctuation differences that occurred for the same responses across tasks (e.g., mean posture - F7) are unclear. One potential reason is different information carried by features. For example, increased skin conductance rate can occur due to both positive (excitement, happy) and negative (anger, anxiety) emotions [35]. This can be a reason for not demonstrating decreased SCR rates (F6) corresponding to the above-discussed increased heart rate variability instance. Another potential reason is the temporal resolution differences in features. For example, heart rate variability (F2) reached its minimums soon after introducing the task and soon after exposing to triggers in many instances, in comparison to other features' corresponding maximums or minimums. Therefore, it can be identified as a fast-responding feature that is suitable for capturing phases early.

In conclusion, it is evident that at least two phases occurred while anticipating and exposing to triggers within anxiety experiences induced in this study. Existing objective anxiety assessment

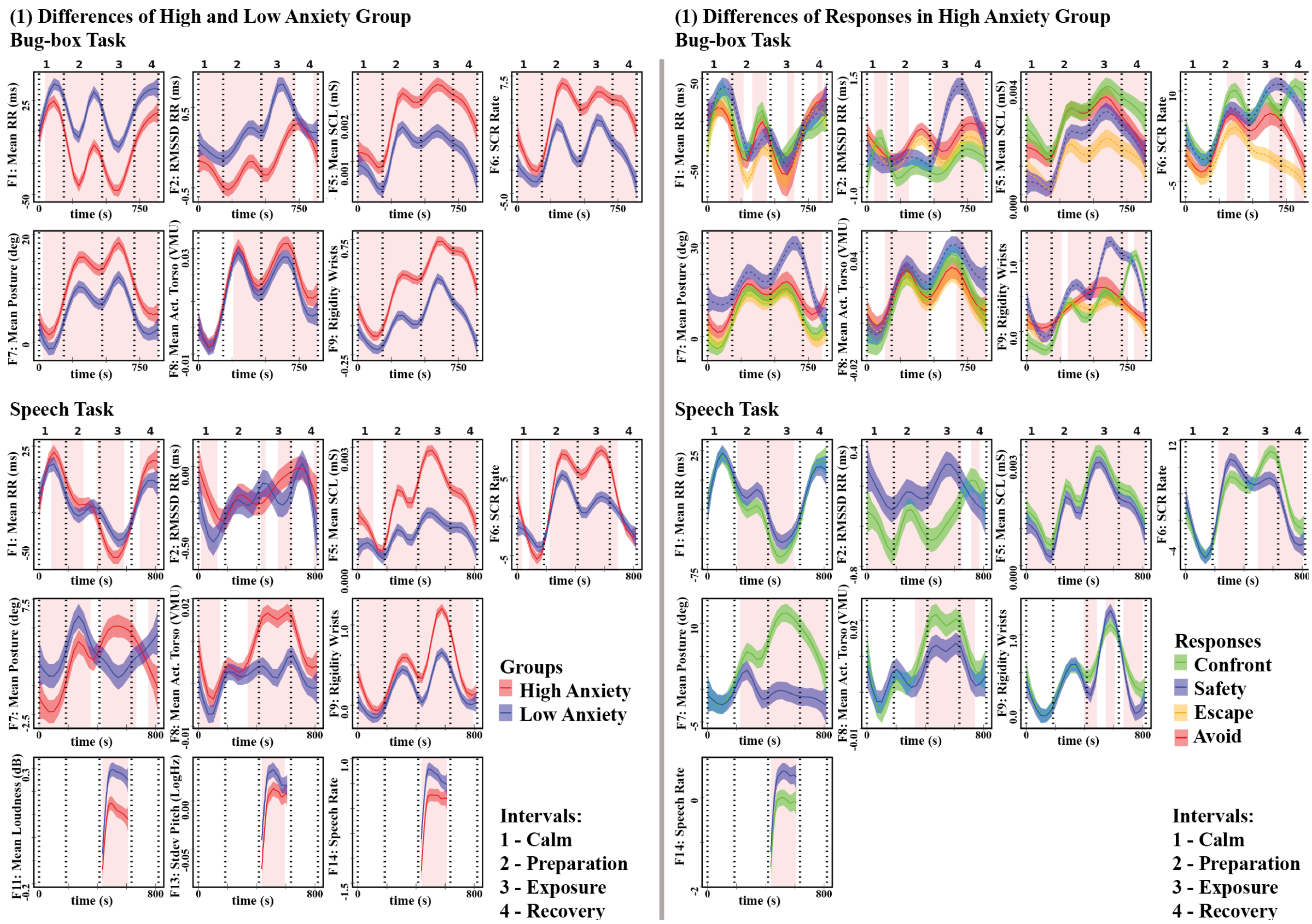

Figure 4: Feature differences from baselines for (1) high and low anxiety groups, and (2) different responses in high anxiety group. Shaded areas around models represent 95% confidence intervals calculated from plot_smooths function, and shaded interval bands show where models significantly differ using plot_difference function (R).

research has an inclination towards analyzing only periods of exposure to triggers [46, 67]. However, appraisals linked to anxiety can arise well before exposures, as also evident in our findings. To this end, future digital interventions can attempt to mitigate the adverse responses (e.g., avoidance, escape) by detecting anticipatory intervals and predicting the upcoming responses. One challenge to the precise detection and prediction would be the physiological suppression likely to occur at adverse responses to certain triggers.

7 LIMITATIONS AND FUTURE WORK

Although our analysis indicated significant differences between reaction-based subgroups, these groups were smaller in sample sizes and did not cover avoidance and escape reactions for speech tasks. Since we accommodated natural reactions to triggers, used convergent ground truths and experienced sensor data losses, this limitation occurred although the overall sample size was relatively large. For the same reason, we could not form corresponding groups with highly matched genders and ethnicity, although participants were within the same age range. In fact, the number of female participants was higher in the HA group, which can also be because anxiety disorders are more prevalent in females [32]. Due to these reasons, we did not include demographic factors as an interaction factor for GAMM models. Future larger studies within controlled settings also need to understand whether the predictive features identified in this study are reasonable to fuse in multimodal-multisensor anxiety phase-detection systems. Further, before moving towards field studies, well-justified feature engineering efforts would be required (which has not been a focus of this study), as many of the features heavily analyzed in state-of-the-art literature are affected by confounding factors (e.g., posture and speaking influence mean RR intervals [12, 40]).

ACKNOWLEDGMENTS

We would like to thank our participants, Mairead Cardamone-Breen and Lachlan James (our support psychologists), and Yasura Vithana (an electronics engineer), for their immense support to the establishment of the described dataset after a Covid-19 lockdown. We thank Julie Holden (Information Technology) for the editorial assistance.

REFERENCES

[1] Ahmet Akin and Bayram Çetin. 2007. The Depression Anxiety and Stress Scale (DASS): The study of Validity and Reliability. *Educational sciences : theory & practice* 7, 1 (2007), 260.

[2] American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders : DSM-5.* (5th ed. ed.). American Psychiatric Pub, Arlington, VA.

[3] David H Barlow. 2001. *Anxiety and its disorders: the nature and treatment of anxiety and panic.* Guilford Press.

[4] Megan Christina Barnsley. 2012. The social consequences of defensive physiological states.

[5] Miguel Barreda-Ángeles, Sara Aleix-Guillaume, and Alexandre Pereda-Baños. 2020. Users' psychophysiological, vocal, and self-reported responses to the apparent attitude of a virtual audience in stereoscopic 360°-video. *Virtual reality : The journal of the Virtual Reality Society* 24, 2 (2020), 289–302.

[6] Aaron T Beck and Emily A.P Haigh. 2014. Advances in Cognitive Theory and Therapy: The Generic Cognitive Model. *Annual review of clinical psychology* 10, 1 (2014), 1–24.

[7] Tom Beckers and Michelle G Craske. 2017. Avoidance and decision making in anxiety: An introduction to the special issue. *Behaviour research and therapy* 96 (2017), 1–2.

[8] Mehdi Boukhechba, Anna N Baglione, and Laura E Barnes. 2020. Leveraging Mobile Sensing and Machine Learning for Personalized Mental Health Care. *Ergonomics in design* 28, 4 (2020), 18–23.

[9] Jason J Braithwaite, Derrick G Watson, Robert Jones, and Mickey Rowe. 2013. A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology* 49, 1 (2013), 1017–1034.

[10] John A Chalmers, James A J Heathers, Maree J Abbott, Andrew H Kemp, and Daniel S Quintana. 2016. Worry is associated with robust reductions in heart rate variability: a transdiagnostic study of anxiety psychopathology. *BMC Psychology* 4, 1 (2016), 32–32.

[11] John A Chalmers, Daniel S Quintana, Maree J-Anne Abbott, and Andrew H Kemp. 2014. Anxiety Disorders are Associated with Reduced Heart Rate Variability: A Meta-Analysis. *Frontiers in psychiatry* 5 (2014), 80–80.

[12] Hsin-An Chang, Chuan-Chia Chang, Nian-Sheng Tzeng, Terry B J Kuo, Ru-Band Lu, and San-Yuan Huang. 2013. Generalized anxiety disorder, comorbid major depression and heart rate variability: a case-control study in taiwan. *Psychiatry investigation* 10, 4 (2013), 326–335.

[13] Nele Dael, Marcello Mortillaro, and Klaus R Scherer. 2012. Emotion Expression in Body Action and Posture. *Emotion (Washington, D.C.)* 12, 5 (2012), 1085–1101.

[14] Damien Dupré, Nicole Andelic, Daniel Stephen Moore, Gawain Morrison, and Gary John McKeown. 2020. Analysis of physiological changes related to emotions during a zipline activity. *Sports engineering* 23, 1 (2020).

[15] T Ehring. 2014. Cognitive Theory. In *The Wiley Handbook of Anxiety Disorders.* John Wiley & Sons, Ltd, Chichester, UK, 104–124.

[16] Mohamed Elgendi and Carlo Menon. 2019. Assessing Anxiety Disorders Using Wearable Devices: Challenges and Future Directions. *Brain sciences* 9, 3 (2019), 50.

[17] A Ellis. 1991. The revised ABC's of rational-emotive therapy (RET). *Journal of rational-emotive and cognitive-behavior therapy* 9, 3 (1991), 139–172.

[18] Norman S Endler and Nancy L Kocovski. 2001. State and trait anxiety revisited. *Journal of anxiety disorders* 15, 3 (2001), 231–245.

[19] Thane M Erickson, Michelle G Newman, and Jamie L Tingey. 2020. Worry and rumination. (2020).

[20] Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on multimedia (MM '10).* ACM, 1459–1462.

[21] Rafael Christophe Freire, Rafael Ferreira-Garcia, Mariana Costa Cabo, Renan Machado Martins, and Antonio Egidio Nardi. 2020. Panic attack provocation in panic disorder patients with a computer simulation. *Journal of affective disorders* 264 (2020), 498–505.

[22] Lior Galili, Ofer Amir, and Eva Gilboa-Schechtman. 2013. Acoustic properties of dominance and request utterances in social anxiety. *Journal of social and clinical psychology* 32, 6 (2013), 651–673.

[23] Denis Gerstorf, Karen L Siedlecki, Elliot M Tucker-Drob, and Timothy A Salthouse. 2009. Within-person variability in state anxiety across adulthood: Magnitude and associations with between-person characteristics. *International journal of behavioral development* 33, 1 (2009), 55–64.

[24] Alexander M Goberman, Stephanie Hughes, and Todd Haydock. 2011. Acoustic characteristics of public speaking: Anxiety and practice effects. *Speech communication* 53, 6 (2011), 867–876.

[25] Dan W Grupe and Jack B Nitschke. 2013. Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nature reviews. Neuroscience* 14, 7 (2013), 488–501.

[26] Ritchie Hannah and Roser Max. 2018. Mental Health. *Our World in Data* (2018). https://ourworldindata.org/mental-health.

[27] Trevor J Hastie and Robert J Tibshirani. 2017. *Generalized additive models.* Routledge.

[28] RG Heimberg, KJ Horner, HR Juster, et al. 1999. Psychometric properties of the Liebowitz Social Anxiety Scale. *Psychological medicine* 29, 1 (1999), 199–212.

[29] Matthias Heyne, Donald Derrick, and Jalal Al-Tamimi. 2019. Native Language Influence on Brass Instrument Performance: An Application of Generalized Additive Mixed Models (GAMMs) to Midsagittal Ultrasound Images of the Tongue. *Frontiers in psychology* 10 (2019), 2597–2597.

[30] Stefan G Hofmann, Stefan M Schulz, Sanna Heering, Frederick Muench, and Lynn F Bufka. 2010. Psychophysiological correlates of generalized anxiety disorder with or without comorbid depression. *International journal of psychophysiology* 78, 1 (2010), 35–41.

[31] Emily M Johnson and Meredith E Coles. 2013. Failure and Delay in Treatment-Seeking Across Anxiety Disorders. *Community mental health journal* 49, 6 (2013), 668–674.

[32] Ronald C Kessler, Maria Petukhova, Nancy A Sampson, Alan M Zaslavsky, and Hans-Ullrich Wittchen. 2012. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *International journal of methods in psychiatric research* 21, 3 (2012), 169–184.

[33] R Kher. 2019. Signal processing techniques for removing noise from ECG signals. *Journal of Biomedical Engineering and Research* 3 (2019), 1–9.

[34] A Konnopka and H Koenig. 2020. Economic Burden of Anxiety Disorders: A Systematic Review and Meta-Analysis. *PharmacoEconomics* 38, 1 (2020), 25–37.

[35] SD Kreibig. 2010. Autonomic nervous system activity in emotion: A review. *Biological psychology* 84, 3 (2010), 394–421.

[36] Martin Lang, Jan Krátký, John H Shaver, Danijela Jerotijević, and Dimitris Xygalatas. 2015. Effects of Anxiety on Spontaneous Ritualized Behavior. *Current biology* 25, 14 (2015), 1892–1897.

[37] Amy E Lawrence, Ronald C Kessler, Dan J Stein, et al. 2008. Assessment of Anxiety Disorders. In *Oxford Handbook of Anxiety and Related Disorders.* Oxford Library of Psychology, Vol. 1. Oxford University Press.

[38] DM Lipnicki and DG Byrne. 2008. An Effect of Posture on Anticipatory Anxiety. *International journal of neuroscience* 118, 2 (2008), 227–237.

[39] Jeongok Logan, SeonAn Yeo, Suk-Sun Kim, and Mijung Lee. 2018. Anxiety and physical inactivity: breaking the vicious circle. *Learning disability practice* 21, 2 (2018), 15.

[40] Valerie A MacIntyre, Peter D MacIntyre, and Geoff Carre. 2010. Heart Rate Variability as a Predictor of Speaking Anxiety. *Communication research reports* 27, 4 (2010), 286–297.

[41] D Makowski, T Pham, Z.J Lau, J.C Brammer, F Lespinasse, H Pham, C Schölzel, and S.H.A Chen. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior research methods* (2021).

[42] Iris Mauss, Frank Wilhelm, and James Gross. 2004. Is there less to social anxiety than meets the eye? Emotion experience, expression, and bodily responding. *Cognition and emotion* 18, 5 (2004), 631–642.

[43] RE McCabe. 2015. Subjective units of distress scale. *Phobias* 18 (2015), 361.

[44] Alicia E Meuret, David Rosenfield, Frank H Wilhelm, Enlu Zhou, Ansgar Conrad, Thomas Ritz, and Walton T Roth. 2011. Do Unexpected Panic Attacks Occur Spontaneously? *Biological psychiatry (1969)* 70, 10 (2011), 985–991.

[45] A Mkrtchian et al. 2017. Modelling avoidance in mood and anxiety disorders using reinforcement-learning. *Biological psychiatry (1969)* 82, 7 (2017), 532–539.

[46] Jennifer Lara Maria Mumm, Lena Pyrkosch, Jens Plag, Patrick Nagel, et al. 2019. Heart rate variability in patients with agoraphobia with or without panic disorder remains stable during CBT but increases following in-vivo exposure. *Journal of anxiety disorders* 64 (2019), 16–23.

[47] Mark Olfson, Mary Guardino, Elmer Struening, Franklin R Schneier, Fred Hellman, and Donald F Klein. 2000. Barriers to the Treatment of Social Anxiety. *The American journal of psychiatry* 157, 4 (2000), 521–527.

[48] Turgut Özseven, Muharrem Düğenci, Ali Doruk, and Hilal I Kahraman. 2018. Voice traces of anxiety: acoustic parameters affected by anxiety disorder. *Archives of Acoustics* (2018), 625–636.

[49] L Parkitny and J McAuley. 2010. The Depression Anxiety Stress Scale (DASS). *Journal of physiotherapy* 56, 3 (2010), 204.

[50] H Posada-Quintero and K Chon. 2020. Innovations in electrodermal activity data collection and signal processing: A systematic review. *Sensors* 20, 2 (2020), 479.

[51] Martin Reuter, Andrew J Cooper, Luke D Smillie, Sebastian Markett, and Christian Montag. 2015. A new measure for the revised reinforcement sensitivity theory: psychometric criteria and genetic validation. *Frontiers in systems neuroscience* 9 (2015), 38–38.

[52] Jan Richter, Alfons O Hamm, Christiane A Pané-Farré, et al. 2012. Dynamics of Defensive Reactivity in Patients with Panic Disorder and Agoraphobia: Implications for the Etiology of Panic Disorder. *Biological psychiatry (1969)* 72, 6 (2012), 512–520.

[53] David Rosenfield, Enlu Zhou, Frank H Wilhelm, Ansgar Conrad, Walton T Roth, and Alicia E Meuret. 2010. Change point analysis for longitudinal physiological data: Detection of cardio-respiratory changes preceding panic attacks. *Biological psychology* 84, 1 (2010), 112–120.

[54] Sainburg, T. 2020. noisereduce 1.1.0: Noise reduction in python using spectral gating. https://pypi.org/project/noisereduce/. Accessed: 2021-05-25.

[55] Seeed Wiki. 2021. *Grove - GSR Sensor.* Seeed Technology Inc., Shenzhen, China.

[56] Hashini Senaratne. 2019. Detecting Temporal Phases of Anxiety in The Wild: Toward Continuously Adaptive Self-Regulation Technologies. In *2019 International Conference on Multimodal Interaction* (Suzhou, China) *(ICMI '19)*. Association for Computing Machinery, New York, NY, USA, 446–452.

[57] Hashini Senaratne, Levin Kuhlmann, Kirsten Ellis, Glenn Melvin, and Sharon Oviatt. 2021. Anxiety Phases Dataset. https://doi.org/10.26180/15176082

[58] F Shaffer and JP Ginsberg. 2017. An Overview of Heart Rate variability Metrics and Norms. *Frontiers in public health* 5 (2017), 258.

[59] Vered Silber-Varod, Hamutal Kreiner, Ronen Lovett, Yossi Levi-Belz, and Noam Amir. 2016. Do social anxiety individuals hesitate more? The prosodic profile of hesitation disfluencies in Social Anxiety Disorder individuals. *Proceedings of Speech Prosody 2016* 31 (2016), 1211–1215.

[60] M Sóskuthy. 2017. Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. *arXiv preprint arXiv:1703.05339* (2017).

[61] LG Tereshchenko and ME Josephson. 2015. Frequency content and characteristics of ventricular conduction. *Journal of electrocardiology* 48, 6 (2015), 933–937.

[62] JF Thayer. 2017. A Neurovisceral Integration Model of Heart Rate Variability. (2017).

[63] Paul van Gent, Haneen Farah, Nicole van Nes, and Bart van Arem. 2019. HeartPy: A novel heart rate algorithm for the analysis of noisy signals. *Transportation research. Part F, Traffic psychology and behaviour* 66 (2019), 368–378.

[64] KJ Wardenaar, CCW Lim, AO Al-Hamzawi, and J Alonso. 2018. The cross-national epidemiology of specific phobia in the World Mental Health Surveys - CORRIGENDUM. *Psychological medicine* 48, 5 (2018), 878–878.

[65] Peter H Waxer. 1977. Nonverbal cues for anxiety: An examination of emotional leakage. *Journal of abnormal psychology (1965)* 86, 3 (1977), 306–314.

[66] Justin W Weeks, Richard G Heimberg, and Reinhardt Heuer. 2011. Exploring the Role of Behavioral Submissiveness in Social Anxiety. *Journal of social and clinical psychology* 30, 3 (2011), 217–249.

[67] Wanhui Wen, Guangyuan Liu, Zhi-Hong Mao, Wenjin Huang, Xu Zhang, Hui Hu, Jiemin Yang, and Wenyan Jia. 2020. Toward Constructing a Real-time Social Anxiety Evaluation System: Exploring Effective Heart Rate Features. *IEEE transactions on affective computing* 11, 1 (2020), 100–110.

[68] WitMotion Team. 2020. *User Manual WT901BLECL BLE5.0: Bluetooth 5.0 Inclinometer Sensor*. WitMotion Shenzhen Co., Ltd., Shenzhen, China.

[69] HU Wittchen, MB Stein, and RC Kessler. 1999. Social fears and social phobia in a community sample of adolescents and young adults: prevalence, risk factors and co-morbidity. *Psychological medicine* 29, 2 (1999), 309–323.

[70] Simon N Wood. 2017. *Generalized additive models: an introduction with R*. CRC press.

[71] World Health Organization. 2017. *Depression and other common mental disorders: global health estimates*. Technical Report.

[72] Zephyr Technology. 2012. *Zephyr™ BioHarness 3.0 User Manual*. Zephyr Technology, Annapolis, MD, United States.