EXPLORING REAL-TIME BIO-BEHAVIORALLY-AWARE FEEDBACK INTERVENTIONS FOR MITIGATING PUBLIC SPEAKING ANXIETY

A Thesis

by

AKANSHA AGARWAL

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Dr. Theodora Chaspari
Committee Members,	Dr. Tracy Hammond
	Dr. Amir Behzadan
Head of Department,	Dr. Scott Schaefer

May 2021

Major Subject: Computer Science

ABSTRACT

Effective public speaking skills are crucial to one's academic and professional success. Individuals who are good in public speaking are more likely to graduate college and obtain a leadership position compared to their counter-peers. Yet, public speaking anxiety (PSA) is one of the most common social fear faced by people directly affecting one's academic and professional success.

This Master's thesis investigates the effectiveness of in-the-moment bio-behaviorally aware feedback in mitigating public speaking anxiety in a virtual training environment. The training environment exposes participants to various virtual stimuli and at the same time, captures their audio and physiological signals. These signals are used to extract bio-behavioral measures (e.g., speech intonation, electrodermal activity mean) and serve as an input to a machine learning model that provides real-time estimates of state anxiety. Based on these state anxiety estimates, the system provides real-time feedback of positive reinforcement and cognitive restructuring–grounded on theoretical rationale from behavioral sciences–when an increase in state anxiety is detected. The system is evaluated through a small-scale study of participants using a pre/post evaluation design.

Results indicate that in-the-moment feedback prompts provided to the participants affect their in-the-moment state-based anxiety. Statistical analysis indicates significant differences of bio-behavioral measures before and after the in-the-moment feedback prompts. The self-reported POST-study results from the participants of the user study also indicate that 5 out of 7 participants found this study beneficial for their public speaking skills. Results of this work also highlight the effect of type of audience on the positive reinforcement feedback provided to the participants. It is observed that when the audience is negative, the positive reinforcement feedback prompts provided to the participants by the real-time model were more compared to a positive audience. Findings from this work provide a foundation toward designing artificial intelligence systems for personalized in-the-moment interventions for mitigating adverse behavioral outcomes.

DEDICATION

To my mother, Jyoti and my father, Punya.

ACKNOWLEDGMENTS

First and foremost I am extremely grateful to my advisor, Dr. Theodora Chaspari for her invaluable advice, constant support, and patience during my thesis study. Her immense knowledge and plentiful experience has encouraged me in all the time of my academic research. I would like to express my deepest appreciation to my committee members, Dr. Tracy Hammond and Dr. Amir Behzadan, for their treasured guidance and insightful feedbacks which brought my work to a higher level.

I would like to give a special thanks to my fellow researchers at HUBBS and CIBER lab, Ehsanul Haque Nirjhar and Md Sakib Nazmus for their valuable assistance and encouragement throughout the research. I am really thankful to my friends Rishabh Garg, Darakshan Anwar, Pranjal Awasthi and Abhay Kumar Singh for their friendship, empathy and encouragement.

Finally, I would like to express my gratitude to my loving parents for their wise counsel and sympathetic ear which helped me in completing my thesis successfully. I would like to thank my brother and sister for being my constant emotional support as well as happy distractions to rest my mind outside of my research. My heartfelt thanks.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

The thesis committee for this work includes Dr. Theodora Chaspari (Chair) and Dr. Tracy Hammond (Member) of the Department of Computer Science at Texas A&M University and Dr. Amir H. Behzadan (Member) of the Department of Construction Science. The user study for this research was conducted with the help of Md Sakib Nazmus, Ehsanul Haque Nirjhar, Dr. Theodora Chaspari and Dr. Amir H. Behzadan. Md Sakib Nazmus took part in the setup of Virtual Reality for user study in this research. All other work conducted as part of this thesis was completed by the author independently.

Funding Sources

This work has been supported by the Engineering information Foundation (EiF) through grant number (18.02). The author gratefully acknowledges the support from EiF. Any opinions, findings, conclusions, and recommendations expressed in this thesis are those of the author and do not necessarily represent those of the EiF.

NOMENCLATURE

PSA	Public Speaking Anxiety
VR	Virtual Reality
ANS	Autonomic Nervous System
EDA	Electrodermal Activity
GSR	Galvanic Skin Response
HRV	Heart Rate Variability
BVP	Blood Volume Pulse
PPG	Photoplethysmogram
BSA	Behavioral Speech Anxiety
CWP	Chest Worn Physiological
WWP	Wrist Worn Physiological
VAD	Voice Activity Detection
SCL	Skin Conductance Level
SCR	Skin Conductance Response
IBI	Inter Beat Interval
HR	Heart Rate

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Effective public speaking skills are crucial to one's academic and professional success. Strong communication skills results in successful project management and implementation and is considered as one of the most important factors for successful collaboration [1]. Individuals with effective public speaking skills are more likely to graduate college and obtain a leadership position compared to their counterpeers [2]. Additionally, good communication skills promote critical thinking and assist in the overall development of an individual [3]. Yet, public speaking anxiety (PSA) is one of the most common social fear faced by people and is more prevalent as compared to other common phobias (e.g., fear of heights, fear of spiders) [4].

Recent studies show that 65% women and 57% men in the United States consider public speaking as their most common fear [5]. A 2014 Forbes survey states that 70% of employees agree to public speaking being crucial to their professional success [6]. Another study shows that individuals with PSA have 15% less chance of obtaining a leadership position, earn 10% less than their peers and are 10% less likely to graduate college [2]. These statistics indicate the importance of such skills in an individual's life.

Although human beings are gifted with the ability to vocalize, knowledge, attitude and skills for effective communication does not come naturally for many people and many times has to be practiced and learned [3]. The biggest cause of anxiety during public speaking lies in the novelty and uncertainty of the assigned task. This can be alleviated by the repeated exposure to public speaking experiences, which can result in gradual change of individuals' negative perception related to public speaking [7].

1.1 Prior Work

Prior work has proposed various visual and tactile stimuli to target the anxiety. Studies in psychological sciences indicate that habituation to the public speaking task through frequent encounter can reduce anxiety by desensitizing the speaker to the public environment [8] and contribute towards restructuring the speaker's negative thoughts [9]. Several studies have explored displaying pictures of social stimuli [10], using imaginary audience to simulate a real life public speaking environment [11], [12] or inviting small-sized real audience [13], [14]. Although effective, these techniques depicted limited performance when compared to immersive and realistic experiences provided by VR environments [15], [16], [17], [18], [19], [20], [21], [22]. Practicing public speaking in a virtual environment was found to be more effective when compared with other techniques requiring imagination, where participants need to visualize their audience [20], [21]. VR environments further allows simulation of real-life large audience in eclectic scenarios [15], [16], [21], which reduces the risk of public embarrassment [21].

Motivated by prior work, this Master's thesis investigates the effectiveness of in-the-moment bio-behaviorally aware feedback in mitigating public speaking anxiety in a virtual training environment. The thesis develops and evaluates a personalized in-the-moment system that provides feedback related to positive reinforcement and cognitive restructuring during public speaking training through a virtual reality environment. The proposed system captures participants' speech and physiological signals in real-time, extracts bio-behavioral features from the recorded data (e.g., speech intonation, electrodermal activity mean), utilizes these features to estimate state-based anxiety, and provides real-time feedback when an increase in this in-the-moment state-based anxiety is detected. The proposed system is evaluated through a small-scale study of participants using a pre/post evaluation design. Specifically, it aims to answer the following research questions :

1. To what extent does bio-behaviorally aware in-the-moment feedback affect participants' bio-behavioral reactions during public speaking?

2. To what extent is the provision of bio-behaviorally aware in-the-moment feedback related to self-reported state- and trait-based anxiety scores?

3. How do different audience reactions (i.e., negative, positive and neutral) affect the provision of bio-behaviorally aware in-the-moment feedback?

Knowledge obtained by this Master thesis will assist in the investigation of a real-time bio-behaviorally-aware feedback model for enhancing virtual reality (VR) training. It further has the potential to provide valuable insights in the design of personalized and effective training mechanisms that can contribute to reducing PSA as well as promoting public speaking skills and public speaking performance.

2. LITERATURE REVIEW AND RESEARCH CONTRIBUTIONS

2.1 Theoretical Models of Public Speaking Anxiety

2.1.1 Three system model of public speaking anxiety

According to the "three system" model, stress caused by situations like public speaking is conveyed in three forms : physiological, cognitive and behavioral [23], [24], [25]. The **physiological** part of the three system model is related to the reactivity of the autonomous nervous system (ANS). The ANS is responsible for the body's response to stress and stress regulation [26]. It has two main parts : sympathetic and parasympathetic system. The sympathetic system is responsible for reaction to stressful situations like threat or injury and is associated with “fight-or-flight" reactions [27]. On the other side, the parasympathetic system controls body functions when the body is at rest and counter balances the sympathetic system. The ANS reactivity is manifested through changes in various body processes, such as blood pressure [28], heart rate [29], sweat [30], etc. These measures are commonly used in research related to PSA, since they can provide useful information about the manifestation of anxiety [29]. The physiological aspect of the PSA can be examined using physiological measures of heart rate and sweat glands, and the emotional arousal can be measured by prosodic changes in the speech [31]. This work utilizes these physiological measures along with speech signals which convey valuable information about a person's confidence, motivation and affective state. Multiple studies indicate that voice features, such as loudness and intonation, are related to stress and help in determining a person's ability to convey their thoughts to the audience [32], [33], [34], [35], [36], [37].

The **cognitive** part of the “three system” model refers to the information collected directly from the speaker performing the public speaking task through self-reports, self-monitoring and/or interviews [38]. This work utilizes the self-reported measures to calculate self-reported scores which serve as ground truth for state-based anxiety. Most of the PSA studies rely on these self-reported measures [23].

Lastly, the **behavioral** aspect is measured as the level of anxiety perceived by the audience in the speaker [39]. This work captures behavioral aspects of PSA through human annotators, who rated the content and context of speech in terms of perceived state-based PSA.

Relevance to this research: The three system model has been utilized less often in the study of PSA. Studies have tried to capture the physiological and cognitive part of the PSA for the detection of anxiety in speakers [31]. This work utilizes all aspects for estimating and quantifying PSA, as well as for evaluating the proposed in-the-moment real-time machine learning model. The physiological aspect is captured via bio-behavioral signals (i.e., physiological and vocal measures) recorded from the speaker. The cognitive aspect is registered via speaker's self-reported indices about their perception on public speaking. Finally, the behavioral aspect is measured via third-party annotations by human experts.

2.1.2 State-trait model of public speaking anxiety

The exploration of behavioral science has shown that PSA can be treated as both a state and a trait condition [8]. Trait anxiety refers to the inherent tendency of an individual to have speaking anxiety whereas state anxiety corresponds to the anxiety experienced during the assigned speaking task. Previous studies have utilized the bio-behavioral signals (e.g., physiological and acoustic features) to predict the state anxiety and then further used the trait-based indices to group the speakers based on their general tendencies [31]. The bio-behavioral measures were combined with individual (e.g., personality, traits) and contextual (e.g., gender, age, native language, ethnicity) features to model the variability of stress across people and conditions [40]. To obtain a better understanding of state based anxiety, the time-based trajectories of PSA variation were studied. The study indicated that time-based representation of bio-behavioral measures was more reliable predictor of stress when compared to corresponding aggregate mean scores [40].

Psychology and communication researches indicate that the relationship between physiological and self-reported measures of PSA depends on a variety of psychological, cognitive and demographic factors [10], [14], [11]. Dimberg *et al.* found that individuals with inherent trait-based public speaking stress were found to depict more reactivity [10]. Well prepared individuals generally demonstrate lower physiological reactivity compared to the speakers who are provided with less time [41]. All these findings indicate a complex relationship among factors contributing to PSA (i.e., physiology, individual traits and contextual factors).

Relevance to this research: This work utilizes the self-reported indices provided by the speakers to determine the in-the-moment state-based anxiety as well as trait-based anxiety scores for the participants of the user study. These scores are used for validation of the real-time bio-behaviorally aware machine learning model.

2.2 Machine learning models of public speaking anxiety

Machine learning is a method of data analysis which can learn, identify patterns and make decisions from minimal human intervention. The applications of machine learning have increased drastically over the past few years [42]. In recent studies, machine learning models are being widely used for detecting anxiety and stress experienced by people all around the world.

The commonly used machine learning models are based on supervised learning methods [43]. Several studies have tried to detect disorders like depression, anxiety and Post Traumatic Stress Disorder (PTSD) using supervised machine learning models like random forest, support vector machine, etc [44] [45] [46] [47]. Some prior work has been done for public speaking anxiety detection using machine learning models. Most of these studies try to automate the public speaking anxiety detection using these models. In [48], ensemble trees were used for automatic assessment of public speaking anxiety and high correlation was achieved between their estimation and ground truth. In [40], Principal Component Analysis (PCA) and k-means clustering were used to design group-based machine learning models which take individual (i.e., trait-based anxiety, personality, etc.) and contextual factors (i.e., age, gender, etc.) into account for detection of public speaking anxiety. Regression algorithms like linear regression and random forest regression have been used to estimate the public speaking anxiety using the bio-behavioral measures along with the individual and contextual factors [31].

Relevance to this research: This work utilizes random forest classifier to obtain real-time state-based anxiety (i.e., for public speaking anxiety) predictions along with in-the-moment bio-behaviorally aware positive reinforcement feedback prompts.

2.3 In-the-moment interventions for mitigating anxiety and supporting performance in high-stake communication tasks

In-the-moment interventions can leverage real-time information from wearable devices and immediately provide the necessary scaffolds and prompts, to encourage and motivate the adoption of good practices during a high-stake communication task.

In-the-moment visual, text, and haptic feedback has been proposed as a method of improving body pose and facial expression during communication tasks, such as public speaking. In the CICERO system, for example, visual feedback has been provided by appropriately modifying audience reactions, as well as using a direct green/red bar indicating performance [32], [34]. In other studies, visual feedback has been given in terms of reconstructing an avatar as the current speaker and providing language prompts [49]. In the Presentation Trainer system, haptic feedback has been provided in the form of vibration from a smart watch, while text feedback contained messages for improving body posture, hand gestures, and voice modulation [36]. Results from these studies are not conclusive; it has been demonstrated in some instances that in-the-moment suggestions are superior to no prompting [36], while in others they have been found to be equivalent in [32], [34].

A limited experimentation on in-the-moment interventions has been mostly performed for well-being (e.g., illness management, alcohol use disorders, sedentary behavior) [50], [51], [52]. In the context of health behavior interventions, the use of mobile technology to deliver in-the-moment support is rooted in theoretical and practical perspectives suggesting that states of vulnerability to adverse health events, as well states as of opportunity for positive changes, can emerge rapidly (e.g., over a few hours, minutes, even seconds) and outside of standard treatment settings [53], [54], therefore timing plays an important role for providing support. In-the-moment feedback has the potential to help individuals reflect in real-time upon their thoughts and emotions, therefore helping them to acknowledge limitations in their train of thought and potential factors that might have contributed to the internalization of feelings of anxiety. Although in-the-moment behavioral support interventions have been considered promising [55], very few studies have evaluated such interventions, mostly because their implementation is quite complex and involves the integration of expertise from multiple fields (e.g., behavioral science, computer science, human-computer interaction). A recent study compared an in-the-moment intervention for stress management (e.g., providing a tailored reminder to use stress management intervention strategies when needed) to a control condition which included random reminders [56]. Results from this study suggest that the group which received in-the-moment reminders reported stressful events less frequently, lower stress severity, less negative affect, lower cortisol levels, less frequent eating, less alcohol consumption, less smoking and better sleep quality compared to the control group. In-the-moment feedback has not been considered in training settings (e.g., public speaking training).

2.4 Research contributions

The main contributions of this research are as follows:

1. In contrast to prior work that estimates a speaker's overall state-based anxiety for the entire public speaking encounter (Section 2.1), this Master's thesis designs machine learning models that estimate state-based anxiety on a moment-to-moment basis. This contributes to administering in-the-moment interventions at the time when the machine learning system detects the participants' anxiety.

2. Given the theoretically-grounded rationale from behavioral sciences and initial indications that in-the-moment interventions might mitigate stress (Session 2.3), this Master's thesis will examine the effect of in-the-moment feedback, administered in the form of positive reinforcement and cognitive restructuring, during public speaking training to alleviate PSA.

3. METHODOLOGY

This chapter describes the methodology used for the development of in-the-moment bio- behaviorally aware machine learning model with positive reinforcement feedback interventions. Section 3.1 outlines the dataset used to design the machine learning system that provides momentary state-based anxiety estimates. Section 3.2 discusses the observational coding that was performed to obtain moment-to-moment annotations of state-based PSA. Finally, Section 3.3 describes the bio-behaviorally-aware in-the-moment training interface, including the design parameters of the machine learning model and training feedback.

3.1 Data description

Here we outline the data that have been used to design the machine learning model of state-based PSA, a more detailed description of which can be found in [31]. Data were collected by a set of 53 participants who went through 10 presentation sessions in a three part experimental procedure (PRE, TEST and POST). The PRE and POST treatment included one session each, in which participants presented in front of a real-life audience. The TEST treatment consisted of 8 sessions in which participants presented in virtual environment. The participants were randomly assigned 8 out of 12 predefined virtual environments with various audience reactions, room conditions, and audience size. For each of these sessions, the participants were given 10 minutes to prepare their presentation based on a randomly assigned topic (e.g., business, history, healthcare, etc.), followed by 5 minutes to speak in front of the audience.

The physiological measures were recorded using the Empatica E4 wrist watch, which provides the following physiological signals :

- Electrodermal activity (EDA): EDA or galvanic skin response (GSR) refers to the changes in human sweat gland activity. These signals help in determination of an individual's emotional state but not the emotion type. Skin conductance levels increase when an individual experiences emotional arousal due to a negative or positive stimuli. The EDA signal consists of a tonic base level driver with slow fluctuations and phasic component with fast variations. When considered together, these provide three EDA metrics, which are as follows :

  - Mean SCL: It is the mean level of the EDA signal.
  - SCR frequency: It is the number of skin conductance responses divided by the duration of the corresponding time segment.
  - Mean SCR amplitude: It is the mean amplitude of skin conductance responses within a time segment

- Blood Volume Pulse (BVP): The BVP signal captures changes in the volume of blood flowing through the arteries and capillaries. These changes correspond to the changes in heart rate and blood flow.
  - Heart rate: It is measured by by detecting peak (beats) from the BVP and computing the lengths of the intervals between adjacent beats.
  - Interbeat Interval (IBI): It is the time between beats and is measures by detecting heartbeat in the BVP input signal.

A microphone was further used to capture the speech signals, from which a total of 7 features are extracted using OpenSMILE toolbox [57]. These audio features are extracted using a 10 second analysis window and are as follows :

- Root Mean Square (RMS) energy: It is the effective value of total acoustic signal waveform (i.e., the area under the curve). In terms of speech, it is the delivered power.
- Fundamental frequency (F0): It is the inverse of the length of pitch period. It indicates how high or low is an individual's voice frequency.
- Zero Crossing Rate (ZCR): This represents the rate of sign change in an acoustic signal.
- Jitter: It is a measure of frequency instability. It is the deviation from true periodicity of a presumably periodic signal.

- Shimmer: It is the measure of amplitude instability.
- Number of pauses: It is a measure of fluency of the speaker.
- Voicing probability: It represents the probability of voice activity which is based on an auto-correlation function.

3.2 Third-party annotations of state-based PSA

The behavioral annotations of perceived state-based anxiety during public speaking were performed using the recorded acoustic signals. The annotator was an undergraduate student in Psychological Science with previous experience in behavioral coding. The annotation was performed using the Noldus Observer XT software [58]. Similar to previous studies [59] [60], each annotator was asked to listen to the recorded audios from the public speaking presentations and rate the perceived state-based anxiety on a 5-point Likert scale. The 5-point Likert scale was selected in accordance to previous methods which have performed similar tasks [59] [60]. In order to ensure consistency and minimize human error, the annotators were asked to listen to each audio once before starting the annotation process, and then go over the annotation process 1-2 times, modifying their annotations if required in order to reduce response delays.

3.3 In-the-moment bio-behaviorally aware machine learning model with positive reinforcement feedback interventions

3.3.1 Data pre-processing

All missing entries of the bio-behavioral measures were substituted by the aggregate value of corresponding feature, computed for each participant. Feature selection is performed using the Spearman's rank correlation between the bio-behavioral measures and the annotations of state-based anxiety. The features which provided decent correlation (i.e. features with correlation of 0.05 and more as well as features with correlation 0f -0.05 and less) with annotations were selected, while the rest were discarded. In an effort to model temporal bio-behavioral changes that might be indicative of state-based anxiety, we further attempted to include the bio-behavioral measures from the previous analysis window as the input to the machine learning model, in addition to the corresponding measures from the current analysis window.

The annotations served as labels for the machine learning model. The 5-point Likert scale annotations were converted to binary annotations using the threshold of level 2. This means that for this study all annotations corresponding the level of 3 or more were considered as the representative of anxiety whereas all annotations on level 2 or below represented absense of anxiety. The threshold of level 2 was determined through experimentation. The imbalance in binary annotations (i.e., the data for anxious state of participants was significantly less than the data for not anxious state) was tackled by oversampling the minority class of the processed dataset and the final dataset obtained after sampling was used in the machine learning model.

3.3.2 Real-time estimation of state-based anxiety

The first component of this work develops a machine learning model that provides real-time estimates of state anxiety during public speaking. The considered model provides an estimate of state anxiety over a pre-defined analysis window of length (W), using the corresponding bio-behavioral measures computed over this time interval. Acoustic features include the root mean square energy, fundamental frequency (F0), number of pauses, jitter, shimmer, zero crossing rate and voicing probability. Physiological features include electrodermal activity (EDA), blood volume pulse (BVP), body temperature and body acceleration. These features comprise the input of machine learning models that predict state anxiety during the public speaking encounter in real-time.

Various types of supervised machine learning algorithms (i.e., linear regression, naive Bayes, bagging, boosting, random forest), feature selection approaches (i.e., correlation of features with annotations), and analysis window lengths (i.e., $W = 10, 15$) were used. The final model used for the study comprised of selected audio and physiological features which included the previous 10 seconds window analysis data. These features, along with the binary annotations, were given as an input to the Random Forest Classifier which provided binary predictions indicating the presence or absence of anxiety within the corresponding time interval.

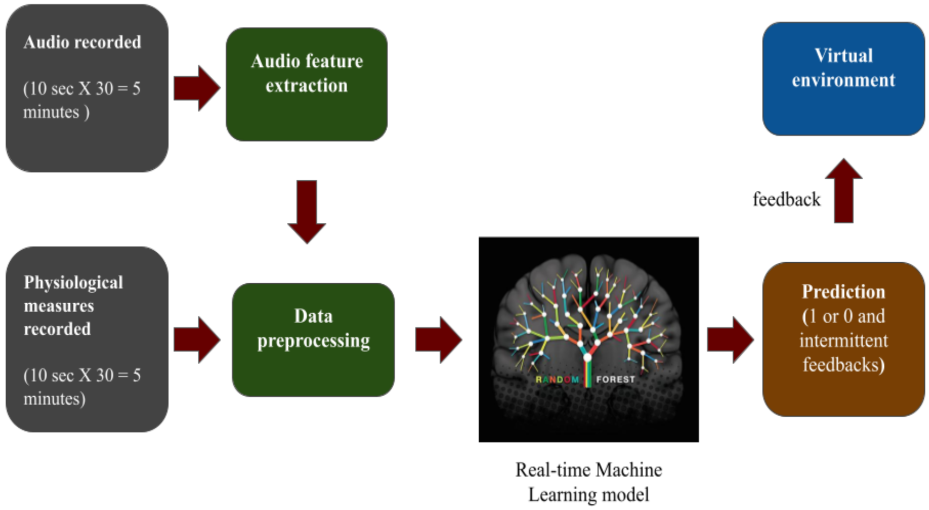

Figure 3.1: Real-time machine learning model with in-the-moment bio-behaviorally aware feedback interventions

In table 3.1, a comparison of various machine learning classifiers is presented. Sampling of data to balance the anxious and non-anxious states along with a 5-fold cross-validation (Leave p-out Cross Validation) was used to generate the accuracy, recall, precision, and F1 scores. The Random Forest classifier was selected as the final classifier as it provided the best results without yielding an excessively high rate of false positive predictions, unlike other classifiers like Bagging and decision tree. Hyper-parameter tuning was performed on Random Forest classifier 3.2, which resulted in a maximum depth of 30 and 30 total trees.

Model	Precision	Recall	F1 score	Accuracy	Max tree depth	# Trees
Logistic Regression	0.25	0.55	0.34	0.53	-	-
Bagging Decision Tree	0.29	0.52	0.37	0.61	30	30
Boosting Decision Tree	0.23	0.36	0.28	0.6	40	40
Random Forest	0.24	0.22	0.23	0.68	30	30

Table 3.1: Comparison of different machine learning models for classifying between the presence or absence of anxiety.

Max tree depth	# Trees	Precision	Recall	F1 score	Accuracy
20	20	0.22	0.24	0.23	0.65
30	30	0.24	0.22	0.23	0.68
40	40	0.25	0.22	0.23	0.68
50	50	0.24	0.22	0.23	0.68

Table 3.2: Hyper-parameter tuning of random forest classifier for classifying between the presence or absence of anxiety.

3.3.3 Design of bio-behaviorally aware in-the-moment feedback

VR interventions, such as cognitive restructuring feedback, have been found to modify speaker's negative perception of a stressful or threatening situation [7], [15], [61], [62]. This work uses feedback, which is determined based on a previous study aimed to understand the preferred mode of feedback by the users when they are receiving public speaking training. The feedback was collected from a survey filled out by the participants of the VR based public speaking user study. Feedback will comprise of the following prompts : "Relax! Take a deep breath!", "You are doing great!", "You know better than the audience", "Don't worry! No one's judging you", "You are doing better than others" and "Stop for 2 seconds and gather your thoughts together". Each participant will be randomly assigned a VR environment from various room conditions (i.e., large hotel room, classroom, meeting room, small theater), audience reactions (i.e., positive, negative, neutral) and audience sizes (i.e., 12, 25, 54, 90), which have also been used in previous studies [31], [61], [18], [63]. The proposed real-time model is an end-to-end model where the participants speak in a virtual environment and are administered with the aforementioned feedback.

4. USER STUDY

The real-time machine learning model with positive reinforcement and cognitive restructuring interventions was evaluated using a small scale user study of 7 participants. Appropriate IRB approval was obtained for this study. Section 4.1 describes the overall structure of the user study. Subsections of 2.1 explains various session divisions within the user study. Section 4.2 lists and details the wearable devices used to capture the various bio-behavioural signals. Finally, section 4.3 explains how the different self-report assessments were used to capture the ground truth state-based anxiety levels of the speakers.

4.1 User study structure

The aim of the user study was evaluation of the real-time machine learning model with positive in-the-moment feedback interventions. The participants were asked to perform public speaking presentations in a virtual environment where they were also given in-the-moment positive reinforcement feedbacks. Participants were graduate students from Texas A&M University and were recruited via emails. All the participants were aged between 22-27 years and were second language speakers. Each participant contributed approximately 2 hours of their time and were provided with a nominal compensation (i.e., $20 Amazon gift card). For this user study each of the 7 participants performed 4 public speaking sessions. The bio-behavioral measures were recorded for a total of 28 sessions, where each session lasted for 5 minutes. Table 4.1 contains further participant details like the ethnicity, age, gender, etc.

4.1.1 Public speaking presentations

The public speaking presentation tasks of the user study consisted of three phases:

- Relaxation phase : This is the first phase of the study. All the participants watched a 5 minute video consisting of nature-based images. The motivation of this phase is to relax and sooth the participant before the actual presentations begin.

individual factors	description
# total participants	7
age range	22-27
average age	24
# males	4
# females	3
ethnicity	6- asian, 1- hispanic/latino
language	all second language speakers
level of education	all graduate students

Table 4.1: Participant details

- Preparation phase : In this phase each participant is provided with 10 minutes to prepare presentation about a randomly assigned topic. There were a total of 30 topics related to history, business, well-being/healthcare, entertainment/culture, technology/science or travel/nature.
- Presentation phase : In the presentation phase, each participant is assigned a random virtual environment in which they speak for 5 minutes.

4.1.2 PRE & POST treatments

Each participant is asked to fill a set of PRE-study and POST-study self-assessment surveys. The PRE-study self-assessments help in determination of trait-based anxiety in a participant. The POST-study self-assessments give information about the state-based anxiety experienced by the participant in the duration of the study.

4.1.3 TEST treatments

The TEST treatment consists of 4 different TEST sessions for each participant. The participants are asked to relax for 5 minutes first by watching a relaxing nature-based video. Each session consists of 10 minutes of preparation time followed by 5 minutes of presentation time. After each presentation, the participants fill TEST surveys which help in determination of state-based anxiety in the participants during the TEST treatments.

Each participant was assigned 4 random virtual environments out of a total of 12 environments. The virtual environments vary in various settings like audience reactions (i.e., negative, positive and neutral), room conditions (i.e., meeting room, classroom and large hotel room) and audience size (i.e., 12, 25, 54) [61], [18], [63], [31].

4.2 Wearable devices

4.2.1 Empatica E4 device : Wrist-worn physiological (WWP) measures

All the participants wore Empatica E4 device which is a wearable research watch used to record real-time physiological measures of the participant. The watch consists of 4 sensors which enable the collection of EDA, BVP, and acceleration data.

4.2.2 Microphone: Acoustic measures from microphone device

During the presentations, all the participants wore a microphone, which helped in recording the audio signals. These audio signals were used to extract the acoustic features using OpenSMILE toolbox [57].

4.2.3 Oculus rift headset & presentation simulator: Virtual environment

Oculus Rift handset [64] and presentation simulator [65] were used to create the virtual environment for the TEST sessions. All the participants were assigned random virtual environments with varied settings of audience reactions (i.e., negative, positive and neutral), room conditions (i.e., meeting room, classroom and large hotel room) and audience size (i.e., 12, 25, 54) [61], [18]. These settings were generated using the presentation simulator and viewed using the Oculus headset. Apart from this, to provide an additional feel of real-life public speaking situation, a low-volume background noise depicting a classroom environment is played on Youtube [66]. This background noise can be heard by the participants through Oculus headphones while they are delivering their oral presentation.

- The **Oculus rift headset** provides an interactive virtual reality experience using 2 screens, one for each eye, which display images side-by-side. A lens set is located at top of the panels which help in focusing, reshaping and creating stereoscopic 3D images for our eye. The Oculus headset has embedded sensors which help in motion recognition and adjustment of the virtual image accordingly [64]. All these features combined help in the creation of a three dimensional virtual reality.

- The **presentation simulator** is a software application designed to be used in Oculus rift. It helps in generation of a virtual audience and public speaking environment for the participants. The features provided by this software allow us to create three types of conference rooms which represent the real-life corporate environment. These rooms offer a variety of audience as well as room sizes. The audience consists of people from various backgrounds (i.e., african, american, asian, etc.) and they have eclectic traits (i.e., men, women, young, old, casual, formal, etc.). The audience shows different kind of reactions which can be categorized into positive, neutral and negative. These behaviors are created such that they mimic different real-life audience circumstances.

4.3 Self-assessments

All the participants were asked to fill in three types of self-assessments : PRE-study, TEST and POST-study. These self-assessments are used to capture the ground truths of the participants. The ground truth corresponds to the self-reported state-based and trait-based anxiety as well as individual and contexual factors like age, gender, ethnicity, etc.

4.3.1 PRE-study self-assessments

These self-assessments were filled by participants before the user study begins.

- State Trait Anxiety Inventory (STAI trait) [67] : STAI is used to measure the state-based and trait-based anxiety of participants. In PRE-study, STAI trait questionnaire is used to record individual's trait-based anxiety scores where higher scores indicate higher anxiety. STAI consists of 20 questions which help in assessment of trait-based anxiety. This self-assessment consists of questions like : "I feel satisfied with myself", "I feel like a failure", "I feel rested", etc, which are rated on a 4-point Likert scale from "Almost Never" to "Almost Always".

- Communication Anxiety Inventory (CAI trait) [68] : CAI is used to measure the state-based and trait-based anxiety of participants. In PRE-study, CAI trait questionnaire is used to record individual's trait-based anxiety scores where higher scores indicate greater anxiety. It consists of 21 questions which are rated on a 4-point Likert scale from "Almost Never" to "Almost Always". This self-assessment consists of questions like : "I enjoy speaking in public.", "I avoid talking with individuals I don't know very well.", "I feel disappointed in myself after speaking in public.", etc, which tell about an individual's communication based trait anxiety. Pre-specified summations of certain selected items from the overall set of 21 questions provide 3 more scores: the CAI dyadic score, CAI Small group score and CAI Public speaking score.

- Personal Report of Public Speaking Anxiety (PRPSA) [69] : PRPSA focuses on measurement of anxiety during public speaking situations rather than general communication based tasks. It consists of 34 questions which are rated on a 5-point Likert scale from "Strongly Disagree" to "Strongly Agree". This self-assessment consists of questions like : "While preparing for giving a speech, I feel tense and nervous.", "I get anxious when I think about a speech coming up.", "I have no fear of giving a speech.", etc, which tell about an individual's presentation based trait anxiety.

- Brief Fear of Negative Evaluation (BFNE) [70] : This survey helps in measuring the fear an individual has for their negative evaluation by the audience. Fear of negative evaluation by the audience is one of the reasons for anxiety faced by speakers in public [71] [72]. This survey attempts to record the feeling of apprehension speakers feel about audience's evaluation as well as the distress they might feel over these negative evaluations. It also tries to record the expectations within speakers for negative evaluation by the audience [73] as well as the sense of dread speakers might feel for getting evaluated unfavourably by the audience. It consists of 12 questions which are rated on a 5-point Likert scale from "Not at all characteristic of me" to "Extremely characteristic of me". This self-assessment consists of questions like : “I often worry that I will say or do the wrong things.”, “If I know someone is judging me, it has little effect on me.”, “I rarely worry about what kind of impression I am making on someone.”, etc, which tell about an individual’s fear or sense of dread with regards to negative evaluation of their speech by others.

- Reticence Willingness to Communicate (RWTC) [74] : This self-assessment tries to measure the speaker’s reluctance towards situations involving communication. It consists of 31 questions which are rated on a 5-point Likert scale from “Strongly Agree” to “Strongly Disagree”. This self-assessment consists of questions like : “In general, I feel at ease when speaking.”, “I like to initiate conversations.”, “I express myself better in speech than in writing.”, etc, which tell about an individual’s disposition towards communicative situations.

- Demographics [40] : The demographics survey helps in capturing the speaker’s age, biological sex, primary language, education, ethnicity, etc. It consists of 11 questions which are rated on different multiple choice-based options. This self-assessment consists of questions like : “What is the degree level that you are currently pursuing?”, “How many times did you have to give a public speech or presentation during the last 3 months?”, “Which college are you currently enrolled in?”, etc, which tells about the individual’s demographics.

- Daily experience [40] : This questionnaire helps in capturing participant’s daily activity which can affect the presentation given by the participant. It consists of 7 questions which are rated on different multiple choice-based options. This self-assessment consists of questions like : “How long ago was your last meal (including breakfast, lunch, dinner)?”, “How many cups of caffeinated drinks (e.g. coffee, tea, red bull) have you consumed today?”, “Has there been a significant event in the past week that could affect your performance in this task?”, etc, which tells about the individual’s activities before they took part in the user study.

4.3.2 TEST self-assessments

These self-assessments are filled by participants after the end of each TEST session.

- State-Anxiety Enthusiasm (SAE) : This self-assessment helps in capturing the state-based anxiety experienced by participants in the preceding TEST sessions where they orally presented in a virtual environment. It consists of 20 questions which which are rated on a 5-point Likert scale from “Strongly Agree” to “Strongly Disagree”. This self-assessment consists of questions like : “I became so excited that I could have continued forever”, “I am not satisfied with my performance.”, “I found it easy to perform the task.”, etc.

- VR Presence (VRP) [75] : It is used to determine the users’ experiences of media as well as how present they were in the interactive virtual environment. This is done by assessing participant’s physiological reactivity. It consists of 8 questions which are rated on a 7-point Likert scale from “Very aware” to “Hardly Aware”. This self-assessment consists of questions like : “I was aware of the real world”, “I wanted to make specific sounds louder or softer”, “I felt I knew what was going to happen next”, etc.

- Presentation Preparation Performance (PPP) [40] : It captures the participants level of performance as well as their level of knowledge about the topic they presented on. It consists of 3 questions which are rated on different multiple choice-based options. This self-assessment consists of the following questions : “How would you rate your level of preparation for the presentation?”, “How would you rate your prior knowledge on the topic that was given to you?” and “How would you rate your performance during the presentation?”.

4.3.3 POST-study self-assessments

These self-assessments are filled by each participant after they are finished with the study (i.e., after the 4 TEST sessions are over).

- State-Anxiety Enthusiasm (SAE) : This is the same self-assessment which was filled by participants after each TEST session’s presentation phase.

- State Trait Anxiety Inventory (STAI state) [67] : STAI is used to measure the state-based and trait-based anxiety of participants. In POST-study, STAI state questionnaire is used to record individual's state-based anxiety scores where higher scores indicate higher anxiety. STAI consists of 20 questions which help in assessment of state-based anxiety. This self-assessment consists of questions like : "I am presently worrying over possible misfortunes ", "I feel frightened ", "I am jittery", etc, which are rated on a 4-point Likert scale from "Not at all" to "Very much so".

- Communication Anxiety Inventory (CAI state) [68] : CAI is used to measure the state-based and trait-based anxiety of participants. In POST-study, CAI state questionnaire is used to record individual's state-based anxiety scores where higher scores indicate greater anxiety. It consists of 20 questions which are rated on a 4-point Likert scale from "Not at all" to "Very much so". This self-assessment consists of questions like : "While talking, I was afraid of making an embarrassing or slip of the tongue.", "I worried about what others thought of me. ", "I could not think clearly when I spoke.", etc, which tell about an individual's communication based state anxiety.
- Body Sensations Questionnaire (BSQ) [76]) : This self-assessment measures speaker's physiological reactivity when they are involved in a public speaking task where higher scores indicate greater physiological reactivity. It consists of 18 questions which are rated on a 5-point Likert scale from "Not at all" to "Extremely". This self-assessment consists of questions like : "Pressure or heavy feeling in chest", "Numbness in arms or legs ", "Dizziness", etc.
- Presentation Preparation Performance (PPP) [40] : This is the same self-assessment which was filled by participants after each TEST session's presentation phase.

4.4 Experimental procedure

Each participant starts with filling up the PRE-study questionnaires which is followed by the relaxation phase. Once the participant is relaxed, 4 TEST sessions are executed in which participant is asked to prepare on a randomly assigned topic for 10 minutes and then perform a public speaking oral presentation in a virtual environment.

The participant is also provided with an in-the-moment positive reinforcement feedback in the same virtual environment in form of presentation slides. The positive feedback used was collected from a survey filled out by the participants of the VR based public speaking user study [40] and included prompts like “Relax! Take a deep breath!", “Stop for 2 seconds and gather your thoughts together", “You are doing great!", “You know better than the audience", “Don’t worry! No one’s judging you", “You are doing better than others" and “Stop for 2 seconds and gather your thoughts together". The presentation slides were automated using python and google slides. The machine learning model provided with a random in-the-moment feedback based on the state-based anxiety prediction by the model. This feedback was displayed in presentation slides within the virtual environment for the speaker to observe. To prevent continuous displaying of these prompts, which can cause disturbance for the speaker, these prompts are programmed such that they are at least 40 seconds apart. The motive of these in-the-moment positive reinforcing interventions was to modify speaker’s negative perception of a stressful or threatening situation [7] (i.e., the public speaking task they are performing in the given virtual environment).

Each of these TEST sessions is followed by TEST self-assessments to measure the state-based anxiety in the participant during the presentation phase. Once all 4 TEST sessions are complete, the participant fills POST-study self-assessment and is compensated with a $20 amazon gift card.

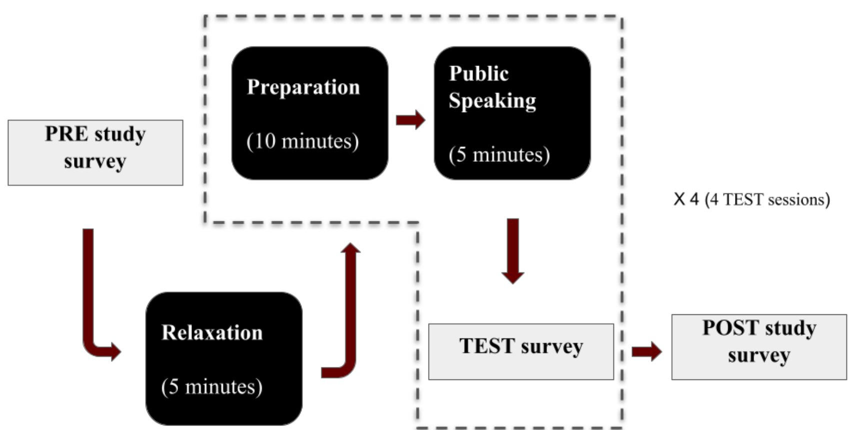

Figure 4.1: User study: The real-time model used for user study

In figure 4.1, the overall structure of user study can be seen. The participants filled PRE-study, POST-study and TEST sessions. The study consisted of three main phases : relaxation, preparation and presentation as shown in the figure.

4.5 Collected data

The bio-behavioral measures, which includes features extracted from physiological and acoustic signals, are collected from this user study. Along with this, data is collected from the PRE-study, TEST and POST-study self-assessment questionnaires which is used to determine participant's in-the-moment state-based and trait-based public speaking anxiety. The in-the-moment positive reinforcement feedback is also collected from the study to analyze the effect of this feedback on participants. Self-assessments also give information about participant's individual and contexual factors. Contextual factors include biological sex, age, native language, ethnicity, highest level of education achieved, current degree, current major, recent public speaking experience, self-reported level of preparation and knowledge on the presentation (PPP). Individual factors include the trait-based anxiety levels (STAI Trait).

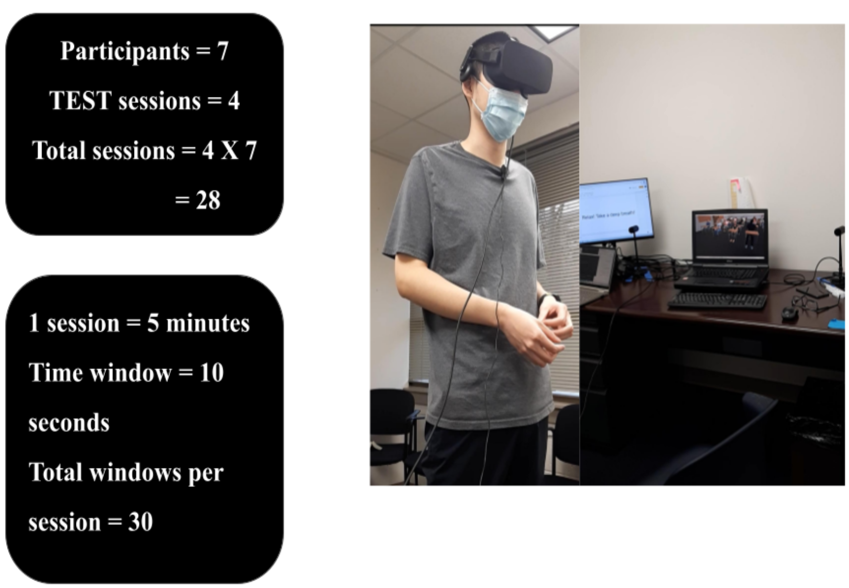

Figure 4.2: User study: A participant speaking in VR

Figure 4.2 shows a participant giving presentation in the assigned virtual environment. The participant is wearing Oculus Rift headset and has microphone attached for recording the acoustic signal.

5. RESULTS

This chapter discusses the analysis performed based on collected data from the user study. We present this analysis in association to the research questions that are formulated as part of this thesis: **(RQ1)** To what extent does bio-behaviorally aware in-the-moment feedback affect participants' bio-behavioral reactions during public speaking? **(RQ2)** To what extent is the provision of bio-behaviorally aware in-the-moment feedback related to self-reported state- and trait-based anxiety scores? **(RQ3)** How do different audience reactions (i.e., negative, positive and neutral) affect the provision of bio-behaviorally aware in-the-moment feedback?

5.1 Effect of in-the-moment feedback prompts on participants' bio-behavioral measures

We investigate the short-term effect of in-the-moment feedback on the participants' bio-behavioral measures by comparing those measures before or after the provision of in-the-moment feedback. Table 5.1 lists all the in-the-moment feedback used in the user study along with the number of times they were displayed in the virtual environment.

In-the-moment feedback	# times used
You are doing great!	15
Relax! Take a deep breath!	13
You are doing better than others	15
You know better than the audience	14
Don't worry! No one's judging you	13
Stop for 2 seconds and gather your thoughts together	21

Table 5.1: In-the-moment feedback used in the virtual environment along with the number of times they were displayed

We compute bio-behavioral measures during the 10 second interval before the provision of feedback, as well as the same measures during the 10 second interval after the feedback provision. We use the Mann-Whitney U test, also known as the Wilcoxon rank sum test [77], a widely used non-parametric method, to test differences with respect to bio-behavioral measures before

Bio-behavioural measures	Mean (±std) (before)	Mean (±std) (after)	Mann Whitney test
Zero crossing rate	0.06 (±0.011)	0.055 (±0.008)	(U=2662.0, p=0.002)
Fundamental frequency	85.96 (±33.7)	88.15 (±30.56)	(U=3441.0, p=0.3)
Voice probability	0.44 (±0.07)	0.45 (±0.06)	(U=3397.0, p=0.25)
RMS energy	0.001 (±0.001)	0.002 (±0.001)	(U=3383.0, p=0.24)

Table 5.2: Mann Whitney U test statistic and p-values for comparison between bio-behavioral measures recorded before the feedback prompts and after these prompts

and after the provision of in-the-moment feedback. In Table 5.2, we observe the results of Mann-Whitney U test statistic as well as the p-value when the bio-behavioral measures recorded before in-the-moment prompts and the bio-behavioral measures recorded after the prompts are compared. Results of the Mann-Whitney test indicate significant differences with respect to these real-time features before and after the in-the-moment feedback interventions. We notice a decrease in zero-crossing rate and an increase in RMS energy of the speech signal, thus indicating that participants' speech rate is decreasing after the feedback provision, while speech loudness is increasing. Similarly, in table 5.3, we observe the results of Mann-Whitney test with respect to specific feedback prompt used during the study. The results in table 5.3 also indicate significant differences between the zero crossing rate before and after the feedback prompt was displayed in the virtual environment. These results also indicate a decrease in speech rate when prompts were given.

5.2 Association between in-the-moment feedback prompts and the VR environment

In response to (RQ2), we examine to what extent in-the-moment feedback prompts are associated with the VR environment, as well as with the participants' anxiety characteristics.

Previous studies have shown that the type of audience is one of the factors affecting state-based anxiety in speaker [17] [31]. We expect that VR environments with negative audience will elicit higher levels of anxiety to participants, therefore the machine learning system will trigger more feedback prompts compared to a VR environment with positive audience. To examine this, we compare the number of prompts that are provided in each VR environment (i.e., positive, neutral, negative). In figure 5.2 the average number of in-the-moment feedback prompts provided to the speakers are shown based on the type of audience. Figure 5.1 shows the total number of in-the-

In-the-moment feedback	Bio-behavioural measures	Mean (±std) (before)	Mean (±std) (after)	Mann Whitney test
Relax! Take a deep breath!	Fundamental frequency	87.3 (± 37.3)	94.9 (± 33.7)	(U= 76.0 , p= 0.34)
	Zero crossing	0.055 (± 0.01)	0.05 (± 0.01)	(U= 53.0 , p= **0.06**)
Stop for 2 seconds and gather your thoughts together	Fundamental frequency	79.03 (± 32.3)	92.8 (± 25.6)	(U= 141.0 , p= **0.06**)
	Zero crossing	0.06 (± 0.01)	0.06 (± 0.01)	(U= 178.0 , p= 0.28)
You are doing great!	Voice probability	0.45 (± 0.05)	0.44 (± 0.07)	(U= 92.0 , p= 0.2)
	Zero crossing	0.061 (± 0.01)	0.05 (± 0.01)	(U= 68.0 , p= **0.03**)
You know better than the audience	Voice probability	0.45 (± 0.07)	0.46 (± 0.07)	(U= 67.0 , p= 0.4)
	Zero crossing	0.054 (± 0.01)	0.05 (± 0.01)	(U= 45.0 , p= **0.06**)
Don't worry! No one's judging you	RMS energy	0.001 (± 0.001)	0.001 (± 0.001)	(U= 66.0 , p= 0.38)
	Zero crossing	0.061 (± 0.02)	0.05 (± 0.01)	(U= 65.0 , p= 0.35)
You are doing better than others	RMS energy	0.0016 (± 0.001)	0.0019 (± 0.001)	(U= 68.0 , p= 0.206)
	Zero crossing	0.065 (± 0.012)	0.056 (± 0.005)	(U= 45.0 , p= 0.023)

Table 5.3: Mann Whitney U test statistic and p-values for comparison between bio-behavioral measures recorded before and after specific feedback prompts

moment feedback prompts provided, as well as the total number of environments used during the user study based on the type of audience. On average, when the audience is positive, less in-the-moment feedback prompts were provided when compared to the negative audience. Table 5.4 shows the average of all state-based in-the-moment anxiety predictions for each audience type. The results indicate that on average, the maximum state-based anxiety was experienced by participants in front of a negative audience while the least amount of state-based anxiety was experienced when the participants were presenting of front of a positive audience, with the neutral audience

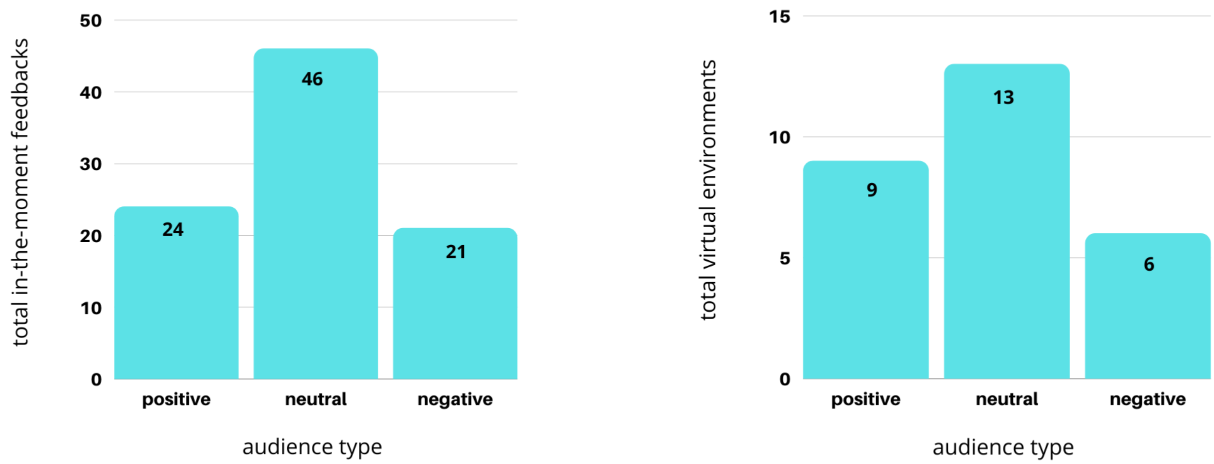

Figure 5.1: Total number of in-the-moment feedback prompts provided during the user study and total number of virtual environments for each type of audience.

Audience type	# Environments	Total predictions	Average predictions
Positive	9	50	5.55
Neutral	13	86	6.61
Negative	6	44	7.33

Table 5.4: Average number of in-the-moment state-based anxiety predictions provided by the real-time model for each type of audience.

giving the average value of 6.61, which is in between the average for positive (i.e., 5.55 average prompts) and negative (i.e., 7.33 average prompts) audience type. In Table 5.6, we can see the results of Mann-Whitney U test [77] to representing pairwise comparisons between the three types of audiences. Even though the dataset is small (negative = 6, positive = 9 and neutral = 13), we observe approaching statistical significance between the negative and positive audience types.

We also compare the three audience types with respect to participants' bio-behavioral measures recorded in each type of audience. In Table 5.5, the mean and standard deviation values of bio-behavioral measures are provided, while Table 5.6 presents the results of Mann-Whitney U test [77] that performs pairwise comparisons between the three types of audiences. Results indicate that participants' heart rate, RMS energy, voice probability, and fundamental frequency are larger

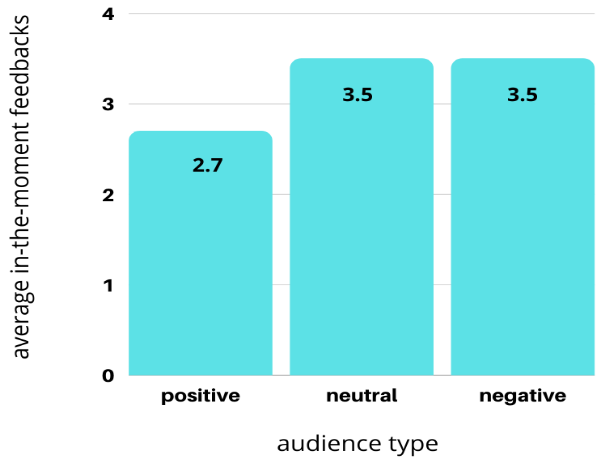

Figure 5.2: Average number of in-the-moment feedback prompts provided to the participants for each type of audience.

in positive audience types, followed by the neutral and the negative audience. High RMS energy in front of the positive audience suggests that participants spoke louder and potentially more confident in these settings, while that was not the case for the negative audience. This difference appears to be statistically significant between the positive and negative audience. Similarly, the higher voice probability in the positive audience potentially means that participants spoke faster due to the fact that they might have felt more comfortable in front of that audience. However, results on the F0 and heart rate are slightly conflicting compared to what we would have expected. These measures are estimates of physiological and vocal reactivity, therefore we would expect those to be higher under the more stressful negative environment. A potential explanation for the large values of F0 and heart rate in the positive environment might be that the participants felt more comfortable in this environment and put increased effort in performing well, therefore yielding an increased cognitive load, which was depicted in higher physiological and vocal reactivity. We also need to consider the additional factors that might be responsible for these results, such as the size of audience, room type, topic provided for public speaking, and participant's traits.

5.3 Association between in-the-moment feedback prompts and participants' anxiety

In response to (RQ3), we anticipate that participants who are inherently more anxious or depict high levels of anxiety during the public speaking task, will receive on average more feedback

Bio-behavioral measure	Audience type	Mean $\pm$ standard deviation
# Prompts	positive	5.55 ± 3.56
	neutral	6.61 ± 4.63
	negative	7.33 ± 2.28
Heart rate	positive	95.74 ± 4.72
	neutral	94.53 ± 4.65
	negative	91.81 ± 3.54
RMS energy	positive	0.0018 ± 0.0014
	neutral	0.0014 ± 0.0011
	negative	0.0007 ± 0.0004
Voice probability	positive	0.46 ± 0.04
	neutral	0.44 ± 0.05
	negative	0.42 ± 0.06
Fundamental frequency (F0)	positive	94.26 ± 26.8
	neutral	84.66 ± 23.8
	negative	73.51 ± 37.12

Table 5.5: Mean and standard deviation for positive, negative and neutral audience type based on number of feedback prompts, heart rate and acoustic measures

prompts from the system. We will quantify anxiety both in terms of the bio-behavioral measures captured during the duration of the public speaking tasks, as well as in terms of the self-reports captured before and after the task, thus obtaining both trait and state anxiety estimates.

Table 5.7 shows the Pearson correlation between bio-behavioral features and number of feedback prompts provided by the system. The observed correlations signify moderate association between the number of prompts and the bio-behavioral measures recorded from the participants. The state-based predictions give correlation with the zero crossing rate, which represents the number of sign changes in a voice signal. Thus, this suggests that the system might have provided more prompts when observing increased speaking rate by the participant. Fundamental frequency, which is the predominant frequency in the recorded acoustic signal, along with root mean square energy of the signal, are also observed to be correlated with the state-based anxiety predictions. Since F0 is a measure of vocal reactivity, which is indicative of state anxiety, it is reasonable that the system provided a larger number of prompts when increased F0 was observed.

The TEST and POST-study self-assessments are used to compute scores which can be used to

Bio-behavioral measure	Audience type 1	Audience type 2	Statistic/p-value
# Prompts	positive	negative	(U=17.0, p=0.13)
	negative	neutral	(U=31.5, p=0.27)
	neutral	positive	(U=62.5, p=0.62)
Heart rate	positive	negative	(U=14.0, p=0.07)
	negative	neutral	(U=25.0, p=0.12)
	neutral	positive	(U=49.0, p=0.27)
RMS energy	positive	negative	(U=10.0, p=0.02)
	negative	neutral	(U=23.0, p=0.09)
	neutral	positive	(U=49.0, p=0.27)
Voice probability	positive	negative	(U=16.0, p=0.11)
	negative	neutral	(U=29.0, p=0.20)
	neutral	positive	(U=41.0, p=0.13)
Fundamental frequency (F0)	positive	negative	(U=19.0, p=0.19)
	negative	neutral	(U=30.0, p=0.23)
	neutral	positive	(U=47.0, p=0.23)

Table 5.6: Mann-Whitney U test for positive, negative and neutral audience type based on number of feedback prompts, heart rate and acoustic measures

Bio-behavioral measure	Pearson's correlation
RMS energy	$r(26) = .12, p = .53$
Zero Crossing	$r(26) = .62, p < 0.001$
Fundamental Frequency	$r(26) = .21, p = .29$

Table 5.7: Pearson's correlation between participants' bio-behavioral measures and number of feedback prompts.

analyse the state-based anxiety predictions of the real-time machine learning model as well as evaluate the significance of recorded in-the-moment bio-behavioral features. After each session, the participants filled three self-assessments: State Anxiety Enthusiasm, VR Sense, and Post Presentation Performance. Table 5.8 shows the Pearson's correlation between self-reported state-based anxiety scores from TEST and the number of in-the-moment prompts provided by the bio-behaviorally aware system. We observe that the system provides a larger number of prompts to participants who reported higher levels of state anxiety. Table 5.9 presents the correlations between the self-reports and the bio-behavioral measures obtained during each session. The positive correlations of the bio-behavioral measures and self-reported measures obtained with the State Anxiety Enthusiasm

TEST self-assessments	Pearson's correlation
State Anxiety Enthusiam	$r(26) = .18, p = .34$
VR Sense	$r(26) = .16, p = .42$
Post Presentation Performance	$r(26) = -.01, p = .96$

Table 5.8: Pearson's correlation between state-based anxiety scores obtained after the end of each presentation and in-the-moment state-based anxiety predictions (# prompts).

TEST self-assessments	bio-behavioral measure	Pearson's correlation
State Anxiety Enthusiam	RMS energy	$r(26) = .22, p = .27$
	Zero Crossing	$r(26) = .29, p = .13$
	Fundamental frequency	$r(26) = -.06, p = .77$
VR Sense	Voice probability	$r(26) = .39, p = .40$
	Zero Crossing	$r(26) = .40, p < 0.05$
	Fundamental frequency	$r(26) = .34, p = .07$
Post Presentation Performance	RMS energy	$r(26) = -.23, p = .24$
	Zero Crossing	$r(26) = -.13, p = .49$
	Voice probability	$r(26) = -.21, p = .27$
	Fundamental frequency	$r(26) = -.18, p = .36$

Table 5.9: Pearson's correlation between state-based anxiety scores obtained after the end of each presentation and the bio-behavioral measures captured during the presentation.

and the VR Sense indicate a positive association between the participants' self-assessments and bio-behavioral measures. Similarly, the negative correlations obtained with the post presentation performance self-assessment scores indicate a negative relationship. The post presentation performance self-assessment is inversely related to the state-based anxiety since a good performance reported by the participant should indicate a decrease in the state-based anxiety experience by the speaker. The observed correlations indicate that the real-time bio-behavioral measures and in-the-moment state-based anxiety predictions are comparable to the state-based anxiety reported by participants in the TEST self-assessments.

Post-study self-assessment scores are used to determine the self-reported measure of state-based anxiety felt by the participant throughout the public speaking user study. The relationship of the overall performance of participants is compared with in-the-moment state-based anxiety predictions received from the real-time machine learning model as well as the bio-behavioral measures extracted from the participants. The post-study self-assessments used in the study are as follows: Post Experimental Feedback (PEF), Post Presentation Performance (PPP), State Anxiety Enthusiasm (SAE), Communication Anxiety Inventory (CAI state), State Trait Anxiety Inventory (STAI state) and Body Sensation Questionnaire (BSQ). Table 5.10 and Table 5.11 shows the Pearson correlation of the calculated state-based anxiety scores from TEST self-assessments with in-the-moment state-based anxiety predictions and the real-time bio-behavioral measures. The negative correlations obtained for the post presentation performance self-assessment scores with RMS energy, voice probability and fundamental frequency indicate a moderate negative relationship between the real-time model and the self-reported POST-study state-based anxiety scores. The post presentation performance self-assessment is inversely related to the state-based anxiety since a good performance reported by the participant should indicate a decrease in the state-based anxiety experience by the speaker. This means that in-the-moment anxiety predictions made by the real-time model are related to the performance reported by the participants in the user study. The positive correlations of the real-time bio-behavioral measures and in-the-moment state-based anxiety predictions obtained with other POST-study self-assessment scores indicates the positive relationship of the calculated scores with the state-based anxiety experienced by participants during the user study.

RMS energy shows correlation of $0.54 (p < 0.005)$ with State Trait Anxiety Inventory self-assessment which means that the energy of the recorded audio signals from speakers' is positively related to the state-based anxiety reported by them. Similarly, the correlation of $-0.56 (p < 0.005)$ observed between RMS energy with Post Presentation Performance self-assessment indicates that this bio-behavioral measure used in the real-time model is strongly related to the performance scores obtained from the self-assessment. Other bio-behavioral measures like zero crossing and voice probability are also observed to have a good relationship with the self-assessment of state-based anxiety by participants. The observed correlations indicate that the real-time bio-behavioral measures and in-the-moment state-based anxiety predictions are comparable to the state-based anxiety reported by participants in the POST self-assessments.

POST self-assessment	Pearson's correlation
State Anxiety Enthusiasm	$r(26) = .1, p = .6$
STAI state	$r(26) = .11, p = .59$
CAI state	$r(26) = .28, p = .15$
Body sensations	$r(26) = .14, p = .49$

Table 5.10: Pearson's correlation of the calculated state-based anxiety scores from POST-study self-assessments with in-the-moment state-based anxiety predictions (# prompts).

POST self-assessment	bio-behavioral measure	Pearson's correlation
State Anxiety Enthusiasm	RMS energy	$r(26) = .42, p = .03$
	zero crossing	$r(26) = .21, p = .29$
Post Experimental Feedback	heart rate	$r(26) = .24, p = .21$
	zero crossing	$r(26) = .11, p = .56$
	voice probability	$r(26) = .56, p = .002$
	fundamental frequency	$r(26) = .4, p = .04$
STAI state	RMS energy	$r(26) = .54, p = .003$
CAI state	RMS energy	$r(26) = .19, p = .34$
	zero crossing	$r(26) = .39, p = .04$
Body sensations	zero crossing	$r(26) = .31, p = .11$
Post Presentation Performance	RMS energy	$r(26) = -.56, p = .002$
	zero crossing	$r(26) = -.18, p = .35$
	voice probability	$r(26) = -.48, p = .01$
	fundamental frequency	$r(26) = -.36, p = .06$

Table 5.11: Pearson's correlation of the calculated state-based anxiety scores from POST-study self-assessments with the real-time bio-behavioral measures.

The trait scores are calculated from pre-study self-assessments. The self-assessments are as follows: Personal Report of Public Speaking Anxiety (PRPSA), Brief Fear of Negative Evaluation (BFNE), Communication Anxiety Inventory (CAI) and State Trait Anxiety Inventory (STAI trait). Table 5.12 shows the Pearson correlation of the calculated trait-based anxiety scores from PRE self-assessments with in-the-moment state-based anxiety predictions and the real-time bio-behavioral measures. As expected, positive associations between the number of prompts and participants' trait anxiety are observed, which means that participants who are inherently more anxius receive a larger number of feedback prompts from the system.

In Table 5.13, we observe moderate positive correlation of various bio-behavioral measures,

PRE self-assessment	Pearson's correlation
BFNE	$r(26) = .24, p = .22$
STAI trait	$r(26) = .16, p = .4$

Table 5.12: Pearson's correlation of the calculated trait-based anxiety scores from PRE-study self-assessments with in-the-moment state-based anxiety predictions (# prompts).

PRE self-assessment	bio-behavioral measure	Pearson's correlation
BFNE	zero crossing	$r(26) = .16, p = .41$
STAI trait	RMS energy	$r(26) = .9, p < .001$
	voice probability	$r(26) = .13, p = .5$
	fundamental frequency	$r(26) = .11, p = .57$
CAI	heart rate	$r(26) = .34, p = .07$
	voice probability	$r(26) = .53, p = .004$
	fundamental frequency	$r(26) = .38, p = .04$
	zero crossing	$r(26) = .23, p = .24$
PRPSA	heart rate	$r(26) = .17, p = .39$
	zero crossing	$r(26) = .25, p = .19$
	voice probability	$r(26) = .29, p = .14$
	fundamental frequency	$r(26) = .29, p = .14$

Table 5.13: Pearson's correlation of the calculated trait-based anxiety scores from PRE-study self-assessments with the real-time bio-behavioral measures.

such as voice probability, fundamental frequency, and RMS energy, with the self-reported trait-based anxiety. These correlations indicate that the trait-based anxiety reported by the participants has a decent relationship with the anxiety predictions made by the real-time model as well as the measures recorded from the participants.

The participants were further asked about their views regarding the study in Post Experimental Feedback self-assessment. The question was as follows: "How much did you feel that this study has benefited your public speaking skills?". The ratings were based on a 5-point Likert scale from "Not at all" to "Extremely". The responses provided by participants are shown in Figure 5.3, where we can observe that some participants felt that this user study helped them improve their public speaking anxiety, while two participants felt that the study did not contribute in improving their public speaking anxiety. A potential reason for this might be the fact that participants only conducted a small number of sessions (i.e., 4). Increasing the number of sessions might have provided

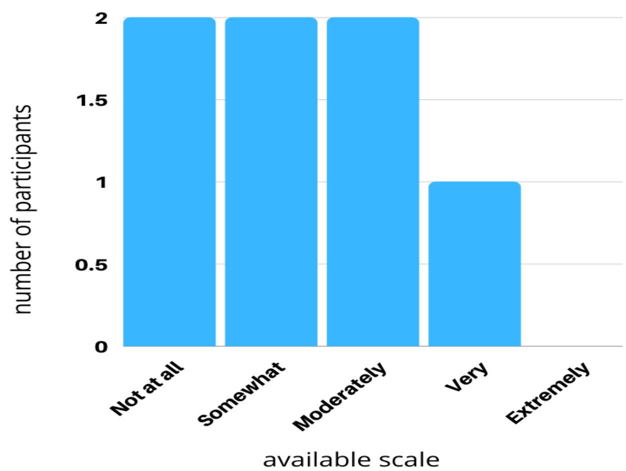

Figure 5.3: Self-reported POST-study assessment results for experienced benefit of the user study

more opportunities to participants in receiving feedback and therefore, might have contributed to helping them with their anxiety.

6. DISCUSSION

The results in this research should be considered in the light of certain limitations. The machine learning model was trained based on annotations obtained from a single experienced annotator. Annotations of the dataset used in study by multiple expert annotators can help in removing the underlying bias, if any. This can help in improving the overall performance of the model as well as in reducing the false positives. There has been limited exploration for in-the-moment feedback intervention methods. Cognitive restructuring feedback is one of the method proposed for tacking the public speaking anxiety. Various other methods of feedback interventions can be explored like vibrations from a wrist watch. One of the limitations of this research is the small dataset used for the analysis, which was conducted with 7 participants. Each participant was part of 4 TEST sessions yielding 28 sessions in total. This dataset is not enough for large-scale analysis and gives comparatively higher p-values, which reduces the statistical significance for some of the results obtained. Only 4 TEST sessions were used in this work compared to 1 PRE, 1 POST and 8 TEST sessions used in preceding work on public speaking anxiety.

The investigation of speaker in virtual environment lacks in certain aspects like analysis of the effect of assigned public speaking topic on participants, number of audience present, and type of room assigned to the participants. In-the-moment feedback prompts provided to participants during their oral presentation in the virtual environment were sometimes not easily visible due to the varying location of presentation slides in the virtual environment. The effect of individual and contextual factors has not been explored in this study. The small size of participants and the lack of diversity in participants prevented this investigation. All participants were graduate level students and were second language speakers. Multiple standard self-assessments were used in the user study to gather information about individual factors, contextual factors, state-based anxiety and trait-based anxiety. These surveys include the assessments about the participant's virtual reality experience but do not properly assess the effect of in-the-moment feedback prompts on the participants. Self-assessments related to in-the-moment feedback can be included in further studies. These assessments can also include a negative scale, in addition to a positive and neutral one, so that accurate assessment of the effect of these prompts on participants public speaking can be captured.

A possible future direction for this work can be exploration of different types of in-the-moment feedback methodologies and their comparison. The study does not include any kind of visual behavioral cues (e.g., facial expression, body gesture), which can be included to improve the performance of in-the-moment model.

7. CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

This research examined the affect of bio-behaviorally aware in-the-moment feedback on the bio-behavioral measures and public speaking anxiety experienced in the virtual environment. Statistical analysis indicates significant differences between bio-behavioral measures recorded before and after the in-the-moment bio-behaviorally aware feedback prompt is displayed to the participants in the virtual environment. Results also indicate that the prompts are provided by the system in reasonable settings. For example, more prompts are provided in negative VR audiences, as well as to participants with higher anxiety levels. A POST-study self-assessment of the participants further suggests that most of the participants found the systematic exposure to VR with in-the-moment reinforcing feedback interventions beneficial for their public speaking skills. The real-time machine learning model with in-the-moment bio-behaviorally aware feedback interventions developed in this research provides promising results which indicate that further exploration of in-the-moment feedback prompts can be beneficial for alleviating public speaking anxiety.

7.2 Future Work

This research explores in-the-moment bio-behaviorally aware feedback interventions and the effect they have on public speaking anxiety in the speaker. The labels of in-the-moment machine learning model developed in this study were based on annotations provided by an experienced annotator. Future studies in this field can use annotations from multiple annotators, which can help in removing the bias and ensure better performance of the model. The presented work also lays foundation for further exploration of different techniques for in-the-moment feedback interventions, such as tactile stimuli, which can assist in real-life interventions to tackle public speaking anxiety.

REFERENCES

[1] G. M. C. PMP, *Communications skills for project managers*. AMACOM Div American Mgmt Assn, 2009.

[2] F. R. Schneier, "Social anxiety disorder," *New England Journal of Medicine*, vol. 355, no. 10, pp. 1029–1036, 2006.

[3] S. P. Morreale, M. M. Osborn, and J. C. Pearson, "Why communication is important: A rationale for the centrality of the study of communication," *JACA-ANNANDALE-*, vol. 1, pp. 1–25, 2000.

[4] R. C. Kessler, M. B. Stein, and P. Berglund, "Social phobia subtypes in the national comorbidity survey," *American Journal of Psychiatry*, vol. 155, no. 5, pp. 613–619, 1998.

[5] K. K. Dwyer and M. M. Davidson, "Is public speaking really more feared than death?," *Communication Research Reports*, vol. 29, no. 2, pp. 99–107, 2012.

[6] C. Gallo, "New survey: 70% say presentation skills are critical for career success. forbes," 2014.

[7] M. T. Triscari, P. Faraci, D. Catalisano, V. D'Angelo, and V. Urso, "Effectiveness of cognitive behavioral therapy integrated with systematic desensitization, cognitive behavioral therapy combined with eye movement desensitization and reprocessing therapy, and cognitive behavioral therapy combined with virtual reality exposure therapy methods in the treatment of flight anxiety: a randomized trial," *Neuropsychiatric disease and treatment*, vol. 11, p. 2591, 2015.

[8] G. D. Bodie, "A racing heart, rattling knees, and ruminative thoughts: Defining, explaining, and treating public speaking anxiety," *Communication education*, vol. 59, no. 1, pp. 70–105, 2010.

[9] J. Ayres and T. Hopf, “Visualization: Reducing speech anxiety and enhancing performance,” *Communication Reports*, vol. 5, no. 1, pp. 1–10, 1992.

[10] U. Dimberg, M. Fredrikson, and O. Lundquist, “Autonomic reactions to social and neutral stimuli in subjects high and low in public speaking fear,” *Biological psychology*, vol. 23, no. 3, pp. 223–233, 1986.

[11] A. Schwerdtfeger, “Predicting autonomic reactivity to public speaking: don’t get fixed on self-report data!,” *International Journal of Psychophysiology*, vol. 52, no. 3, pp. 217–224, 2004.

[12] G. Erdmann and S. Baumann, “Are psychophysiologic changes in the" public speaking" paradigm an expression of emotional stress?,” *Zeitschrift fur experimentelle Psychologie: Organ der Deutschen Gesellschaft fur Psychologie*, vol. 43, no. 2, pp. 224–255, 1996.

[13] A. W. Zuardi, J. A. d. S. Crippa, J. E. C. Hallak, and R. Gorayeb, “Human experimental anxiety: actual public speaking induces more intense physiological responses than simulated public speaking,” *Brazilian Journal of Psychiatry*, vol. 35, no. 3, pp. 248–253, 2013.

[14] C. Kirschbaum, J. C. Prussner, A. A. Stone, I. Federenko, J. Gaab, D. Lintz, N. Schommer, and D. H. Hellhammer, “Persistent high cortisol responses to repeated psychological stress in a subpopulation of healthy men,” *Psychosomatic medicine*, vol. 57, no. 5, pp. 468–474, 1995.

[15] P. L. Anderson, E. Zimand, L. F. Hodges, and B. O. Rothbaum, “Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure,” *Depression and anxiety*, vol. 22, no. 3, pp. 156–158, 2005.

[16] S. R. Harris, R. L. Kemmerling, and M. M. North, “Brief virtual reality therapy for public speaking anxiety,” *Cyberpsychology & behavior*, vol. 5, no. 6, pp. 543–550, 2002.

[17] D.-P. Pertaub, M. Slater, and C. Barker, “An experiment on public speaking anxiety in response to three different types of virtual audience,” *Presence: Teleoperators & Virtual Environments*, vol. 11, no. 1, pp. 68–78, 2002.

[18] J. Diemer, A. Mühlberger, P. Pauli, and P. Zwanzger, "Virtual reality exposure in anxiety disorders: Impact on psychophysiological reactivity," *The World Journal of Biological Psychiatry*, vol. 15, no. 6, pp. 427–442, 2014.

[19] M. Slater, D.-P. Pertaub, and A. Steed, "Public speaking in virtual reality: Facing an audience of avatars," *IEEE Computer Graphics and Applications*, vol. 19, no. 2, pp. 6–9, 1999.

[20] B. L. Heuett and K. B. Heuett, "Virtual reality therapy: A means of reducing public speaking anxiety," *International journal of humanities and social science*, vol. 1, no. 16, pp. 1–6, 2011.

[21] G. Riva, "Virtual reality in neuro-psycho-physiology: Cognitive, clinical and methodological issues in assessment and rehabilitation," 1997.

[22] I. L. Kampmann, P. M. Emmelkamp, D. Hartanto, W.-P. Brinkman, B. J. Zijlstra, and N. Morina, "Exposure to virtual social interactions in the treatment of social anxiety disorder: A randomized controlled trial," *Behaviour Research and Therapy*, vol. 77, pp. 147–156, 2016.

[23] J. C. McCroskey, "The communication apprehension perspective," *Avoiding communication: Shyness, reticence, and communication apprehension*, pp. 13–38, 1984.

[24] M. L. Patterson and V. Ritts, "Social and communicative anxiety: A review and meta-analysis," *Annals of the International Communication Association*, vol. 20, no. 1, pp. 263–303, 1997.

[25] V. P. Richmond and M. Hickson, *Going public: A practical guide to public talk*. Allyn and Bacon, 2002.

[26] J. Andreassi, "Heart activity and behavior ii: Stress, emotions, motivation, personality, social factors, brain interactions, and conditioning," *Psychophysiology: Human behavior and physiological response (5 edn). Mahwal, NJ: Lawrence Erlbaum Associates, Inc., Publishers*, 2007.

[27] A. Sallinen-Kuparinen, J. C. McCroskey, and V. P. Richmond, "Willingness to communicate, communication apprehension, introversion, and self-reported communication competence:

Finnish and american comparisons," *Communication Research Reports*, vol. 8, no. 1, pp. 55–64, 1991.

[28] M. Dickens and W. R. Parker, "An experimental study of certain physiological, introspective and rating-scale techniques for the measurement of stage fright," *Communications Monographs*, vol. 18, no. 4, pp. 251–259, 1951.

[29] R. R. Behnke and L. W. Carlile, "Heart rate as an index of speech anxiety," 1971.

[30] K. Clements and G. Turpin, "Physiological effects of public speaking assessed using a measure of palmar sweating.," *Journal of Psychophysiology*, 1996.

[31] M. Yadav, M. N. Sakib, E. H. Nirjhar, K. Feng, A. Behzadan, and T. Chaspari, "Exploring individual differences of public speaking anxiety in real-life and virtual presentations," *IEEE Transactions on Affective Computing*, 2020.

[32] L. Batrinca, G. Stratou, A. Shapiro, L.-P. Morency, and S. Scherer, "Cicero-towards a multimodal virtual audience platform for public speaking training," in *International workshop on intelligent virtual agents*, pp. 116–128, Springer, 2013.

[33] L. Chen, G. Feng, J. Joe, C. W. Leong, C. Kitchen, and C. M. Lee, "Towards automated assessment of public speaking skills using multimodal cues," in *Proceedings of the 16th International Conference on Multimodal Interaction*, pp. 200–203, 2014.

[34] M. Chollet, T. Wörtwein, L.-P. Morency, A. Shapiro, and S. Scherer, "Exploring feedback strategies to improve public speaking: an interactive virtual audience framework," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 1143–1154, 2015.

[35] M. N. Bos, "Automatic descriptive scoring of public speaking performance based on prosodic features," 2017.

[36] J. Schneider, D. Börner, P. Van Rosmalen, and M. Specht, "Presentation trainer, your public speaking multimodal coach," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pp. 539–546, 2015.

[37] A. C. Aguiar, M. Kaiseler, H. Meinedo, T. E. Abrudan, and P. R. Almeida, "Speech stress assessment using physiological and psychological measures," in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pp. 921–930, 2013.

[38] L. E. Michelson and L. Ascher, *Anxiety and stress disorders: Cognitive-behavioral assessment and treatment.* Guilford Press, 1987.

[39] A. Mulac and A. Sherman, "Conceptual foundations of the behavioral assessment of speech anxiety," *Western Journal of Speech Communication*, vol. 39, pp. 176–180, 1975.

[40] M. Yadav, M. N. Sakib, K. Feng, T. Chaspari, and A. Behzadan, "Virtual reality interfaces and population-specific models to mitigate public speaking anxiety," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 1–7, IEEE, 2019.

[41] M. J. Beatty and R. R. Behnke, "Effects of public speaking trait anxiety and intensity of speaking task on heart rate during performance," *Human Communication Research*, vol. 18, no. 2, pp. 147–176, 1991.

[42] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[43] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

[44] A. Priya, S. Garg, and N. P. Tigga, "Predicting anxiety, depression and stress in modern life using machine learning algorithms," *Procedia Computer Science*, vol. 167, pp. 1258–1267, 2020.

[45] A. G. Reece, A. J. Reagan, K. L. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, "Forecasting the onset and course of mental illness with twitter data," *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.

[46] S. R. Braithwaite, C. Giraud-Carrier, J. West, M. D. Barnes, and C. L. Hanson, "Validating machine learning algorithms for twitter data against established measures of suicidality," *JMIR mental health*, vol. 3, no. 2, p. e21, 2016.

[47] J. Du, Y. Zhang, J. Luo, Y. Jia, Q. Wei, C. Tao, and H. Xu, "Extracting psychiatric stressors for suicide from social media using deep learning," *BMC medical informatics and decision making*, vol. 18, no. 2, pp. 77–87, 2018.

[48] T. Wörtwein, L.-P. Morency, and S. Scherer, "Automatic assessment and analysis of public speaking anxiety: A virtual audience case study," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 187–193, IEEE, 2015.

[49] F. Dermody and A. Sutherland, "A multimodal system for public speaking with real time feedback," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pp. 369–370, 2015.

[50] D. Ben-Zeev, C. J. Brenner, M. Begale, J. Duffecy, D. C. Mohr, and K. T. Mueser, "Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia," *Schizophrenia bulletin*, vol. 40, no. 6, pp. 1244–1253, 2014.

[51] S. Van Dantzig, G. Geleijnse, and A. T. Van Halteren, "Toward a persuasive mobile application to reduce sedentary behavior," *Personal and ubiquitous computing*, vol. 17, no. 6, pp. 1237–1246, 2013.

[52] D. H. Gustafson, F. M. McTavish, M.-Y. Chih, A. K. Atwood, R. A. Johnson, M. G. Boyle, M. S. Levy, H. Driscoll, S. M. Chisholm, L. Dillenburg, *et al.*, "A smartphone application to support recovery from alcoholism: a randomized clinical trial," *JAMA psychiatry*, vol. 71, no. 5, pp. 566–572, 2014.

[53] C. Abraham and S. Michie, "A taxonomy of behavior change techniques used in interventions.," *Health psychology*, vol. 27, no. 3, p. 379, 2008.

[54] T. Webb, J. Joseph, L. Yardley, and S. Michie, "Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of

behavior change techniques, and mode of delivery on efficacy," *Journal of medical Internet research*, vol. 12, no. 1, p. e1376, 2010.

[55] I. Nahum-Shani, S. N. Smith, B. J. Spring, L. M. Collins, K. Witkiewitz, A. Tewari, and S. A. Murphy, "Just-in-time adaptive interventions (jitais) in mobile health: key components and design principles for ongoing health behavior support," *Annals of Behavioral Medicine*, vol. 52, no. 6, pp. 446–462, 2018.

[56] J. M. Smyth and K. E. Heron, "Is providing mobile interventions" just-in-time" helpful? an experimental proof of concept study of just-in-time intervention for stress management," in *2016 IEEE Wireless Health (WH)*, pp. 1–7, IEEE, 2016.

[57] F. Eyben, F. Weninger, M. Wöllmer, and B. Shuller, "Open-source media interpretation by large feature-space extraction," *TU Munchen, MMK*, 2016.

[58] P. H. Zimmerman, J. E. Bolhuis, A. Willemsen, E. S. Meyer, and L. P. Noldus, "The observer xt: A tool for the integration and synchronization of multimodal signals," *Behavior research methods*, vol. 41, no. 3, pp. 731–735, 2009.

[59] C. Busso, S. Parthasarathy, A. Burmania, M. AbdelWahab, N. Sadoughi, and E. M. Provost, "Msp-improv: An acted corpus of dyadic interactions to study emotion perception," *IEEE Transactions on Affective Computing*, vol. 8, no. 1, pp. 67–80, 2016.

[60] I. Lefter, G. J. Burghouts, and L. J. Rothkrantz, "An audio-visual dataset of human–human interactions in stressful situations," *Journal on Multimodal User Interfaces*, vol. 8, no. 1, pp. 29–41, 2014.

[61] D.-P. Pertaub, M. Slater, and C. Barker, "An experiment on public speaking anxiety in response to three different types of virtual audience," *Presence: Teleoperators & Virtual Environments*, vol. 11, no. 1, pp. 68–78, 2002.

[62] D. A. Clark, "Cognitive restructuring," *The Wiley handbook of cognitive behavioral therapy*, pp. 1–22, 2013.

[63] M. N. Sakib, M. Yadav, T. Chaspari, and A. H. Behzadan, "Coupling virtual reality and physiological markers to improve public speaking performance," in *International Conference on Construction Applications of Virtual Reality*, 2019.

[64] "Oculus rift." `https://www.oculus.com/?locale=en_US`. Accessed: 2021-02-23.

[65] "Presentation simulator." `http://www.presentationsimulator.com`. Accessed: 2021-02-23.

[66] "Background noise." `https://www.youtube.com/watch?v=ApYyTBsn2K0&t=1929s`. Accessed: 2021-02-23.

[67] C. Spielberger, "State-trait anxiety inventory. the corsini encyclopedia of psychology," *Hoboken: Wiley*, vol. 1, 2010.

[68] S. Booth-Butterfield and M. Gould, "The communication anxiety inventory: Validation of state-and context-communication apprehension," *Communication Quarterly*, vol. 34, no. 2, pp. 194–205, 1986.

[69] J. C. McCroskey, "Measures of communication-bound anxiety," 1970.

[70] A. Tavoli, M. Melyani, M. Bakhtiari, G. H. Ghaedi, and A. Montazeri, "The brief fear of negative evaluation scale (bfne): translation and validation study of the iranian version," *BMC psychiatry*, vol. 9, no. 1, pp. 1–6, 2009.

[71] D. M. Clark, W. Crozier, and L. Alden, "A cognitive perspective on social phobia," *The essential handbook of social anxiety for clinicians*, pp. 193–218, 2005.

[72] R. M. Rapee and R. G. Heimberg, "A cognitive-behavioral model of anxiety in social phobia," *Behaviour research and therapy*, vol. 35, no. 8, pp. 741–756, 1997.

[73] D. Watson and R. Friend, "Measurement of social-evaluative anxiety.," *Journal of consulting and clinical psychology*, vol. 33, no. 4, p. 448, 1969.

[74] M. Pörhölä, "Trait anxiety, experience, and the public speaking state responses of finnish university students," *Communication research reports*, vol. 14, no. 3, pp. 367–384, 1997.

[75] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[76] D. L. Chambless, G. C. Caputo, P. Bright, and R. Gallagher, "Assessment of fear of fear in agoraphobics: the body sensations questionnaire and the agoraphobic cognitions questionnaire.," *Journal of consulting and clinical psychology*, vol. 52, no. 6, p. 1090, 1984.

[77] P. E. McKnight and J. Najab, "Mann-whitney u test," *The Corsini encyclopedia of psychology*, pp. 1–1, 2010.