Virtual reality interfaces and population-specific models to mitigate public speaking anxiety

Megha Yadav[1], Md Nazmus Sakib[2], Kexin Feng[1], Amir Behzadan[2], Theodora Chaspari[1]
[1]Human Bio-Behavioral Signals (HUBBS) Lab, Texas A&M University, College Station, TX, USA
[2]Construction Informatics and Built Environment Research (CIBER) Lab, Texas A&M University, College Station, TX, USA
{meghakyadav,mnsakib,kexin0814,abehzadan,chaspari}@tamu.edu

***Abstract*—Public speaking is key to effectively exchanging ideas, persuading others, and making tangible impact. Yet, public speaking anxiety (PSA) ranks as a top social phobia among many people. This paper leverages bio-behavioural indices captured from wearable devices and virtual reality (VR) interfaces to quantify PSA. The significance of individual-specific factors, such as general trait anxiety and personality, as well as contextual factors, such as age, gender, highest education, and native language, in moderating the association between bio-behavioral indices and PSA is further examined through group-based machine learning models. Results highlight the importance of including such factors for detecting PSA with the proposed group-based PSA models yielding Spearman's correlation of $0.55 (p < 0.05)$ between the actual and predicted state-based anxiety scores. This work further analyzes whether systematic exposure to public speaking tasks in the VR environment can help alleviate PSA. Results indicate that systematic exposure to public speaking in VR can alleviate PSA in terms of both self-reported ($p < 0.05$) and physiological ($p < 0.05$) indices. Findings of this study will enable researchers to better understand antedecedents and causes of PSA contributing to behavioral interventions using VR.**

***Index Terms*—public speaking anxiety, virtual reality, physiological signals, speech, wearable devices, group-based clustering**

I. Introduction

Public speaking skills are essential to help people effectively exchange ideas [1] and comprise a major factor of one's academic and professional success [2]. Yet, public speaking anxiety (PSA) ranks as a top social phobia in the U.S. [3] and tends to be aggravated in minorities, first generation students, and non-native speakers [4]. PSA is a communication-based disorder that involves the experience of physiological arousal, negative cognition, and behavioral reactions in response to a real or anticipated enactment of oral presentation [5].

Recent advances in wearable devices provide a unique opportunity to explore PSA in various naturalistic settings and situations. Previous research has employed self-reported and signal-based measures to quantify PSA. The former refers to the speaker's own views obtained through self-assessments, while the latter includes physiological responses of the autonomic nervous system (e.g., cardiovascular and electrodermal activity), speech intonation, facial expressions, and body gestures [6], [7]. Preliminary studies have explored several ways to elicit PSA. These include showing pictures of social stimuli (e.g., faces) [8], instructing speech delivery to an imaginary audience [9], [10], or presenting in front of a small-size real audience [11], [12]. With the advent of virtual interfaces, barriers related to providing naturalistic public speaking stimuli have been reduced, with a variety of recent studies exploring the feasibility of virtual reality (VR) applications for studying and quantifying public speaking skills, performance, and anxiety [13]–[19]. VR offers a potential remedy to achieving this through the immersive experience of presenting in various public speaking stimuli without the risk of public embarrassment [20], [21].

In contrast to public speaking performance, PSA is a complicated psychological phenomenon confounded by various individual and contextual factors [8], [10], [11]. Individuals with high trait-based anxiety depict higher physiological reactivity to the public speaking stimuli compared to their low trait anxiety peers [8]. Female speakers depict increased self-reported and physiological anxiety when speaking in public compared to their male counter-parts [22]. Well-prepared individuals have lower physiological reactivity compared to those who have spent less time over preparation [23].

The primary objective of this work is to (1) study the interplay between bio-behavioral indices (physiological, acoustic) and individual and contextual factors to quantify PSA; and (2) analyze whether systematic exposure through VR can help alleviate PSA. To this end, physiological measures of electrodermal activity (EDA), blood volume pulse (BVP), electrocardiogram (ECG), body temperature, body acceleration, and speech were collected during public speaking presentations. The bio-behavioral indices from these signals are studied in association to retrospective self-reported state-based PSA. Individual-specific indices (e.g., demographics) and contextual factors (e.g., frequency of engaging in public speaking, degree of preparation) are incorporated as moderating factors between bio-behavioral indices and state-based anxiety while constructing the PSA models. Our results yield Spearman's correlation of 0.55 (p <0.05) for estimating PSA, and indicate that VR interventions can alleviate participants' PSA, , as measured with subjective self-reports ($t(26)$=2.69; p<0.05) and signal-based bio-behavioral indices ($t(26)$=3.33; p<0.05).

II. Previous Work

Studies in psychology and communication indicate that PSA can be reduced via systematic exposure to public speaking encounters, which can potentially lead to the desensitization of threatening stimuli [5]. Because of its immersiveness, VR can simulate types of public speaking difficult to replicate in real-life [13], [14], [20], [21], [24]. Indicatively, Pertaub et al. [21], found that individuals experience significantly high anxiety during the exposure to negative VR audiences. North et al. [20], reasoned that VR can help individuals who have difficulty imagining public speaking scenarios. Harris et al. [14], reported that a set of four VR sessions can reduce PSA. While previous studies in life sciences have measured PSA through self-reported and physiological indices, this work assesses the effectiveness of VR through a multimodal set of bio-behavioral indices related to speech and physiology.

Previous studies in Affective Computing have used visual and haptic feedback in order to improve public speaking skills. In Cicero, Chollet et al. [25], [26], proposed a 2D avatar augmented with visual stimuli, as provided through a color-coded visual feedback or through the interaction with the virtual audience. In the same study, public speaking performance was quantified through a set of multimodal indices related to speech, vision, and physiology. In Presentation Trainer, Schneider et al. [2], did not use an audience, but provided feedback to the user through his/her mirrored image combined with visual and haptic stimuli.

Previous research further indicates that the association between physiological and state-based measures of PSA is moderated by psychological, demographic, and situational factors. Dimberg et al. [8] found that individuals with high trait-based public speaking fear, depicted increased physiological reactivity. Kirschbaum et al. [11] suggested the presence of two groups of individuals (low and high responders) formed based on personality characteristics. Other studies suggest that physiological reactivity during moments of anxiety is further moderated by the knowledge of the presentation topic, novelty of the presentation, and audience reaction [15], [23]. There is a general consensus in the previous work that a complex interplay exists between between bio-behavioral measures and individual and contextual factors that contribute to PSA.

The main contributions of the research presented in this paper to the body of knowledge lies in the following: 1) Studies in Affective Computing focus on public speaking performed in front of a 2D audience in Cicero [6], [25], [26], or no audience in Presentation Trainer [2], therefore potentially lacking user immersion, which can be provided by the VR; 2) With the exception of [26], previous studies are heavily focused on improving public speaking skills and performance. Yet, PSA is a complicated psychological phenomenon affected by various individual and contextual factors. Taking these factors into account is crucial for modeling PSA and intervening upon it; 3) Previous studies have not taken into account the various individual and contextual factors to quantify PSA. The current paper integrates these factors into group-specific machine learning models that can more accurately estimate PSA compared to general models.

III. Data Description

Our data comes from a user study including public speaking presentations in front of real-life and virtual audiences (Table I). Each participant went through 10 presentation sessions during the three parts of the experimental procedue (PRE, TEST, POST). The PRE and POST treatments each lasted one session, during which participants had to present in front of a real-life audience in order to assess pre- and post-differences.. The TEST treatment consisted of 8 sessions, distributed across 2 days, during which participants wore an Oculus Rift headset and present in front of different groups of VR audiences generated in the Presentation Simulator software [27]. Each participant was randomly assigned to 8 out of 12 VR settings from various room conditions, audience reactions, and audience size [15], [24]. During each session, the participants were given 10 minutes to prepare a presentation based on a randomly assigned news article from various topics (e.g. history, business, well-being/healthcare), and followed by another 5 minutes to present in front of real (PRE, POST) or VR (TEST) audiences.

TABLE I
DATA COLLECTION SETTINGS

	PRE	**TEST**	**POST**
Audience	Real	Virtual	Real
# Sessions	1	8	1
# Participants	55	38	29
# Female	23	16	13
Average age	21	21	21

Participants wore the wrist-mounted Empatica E4 [28], which captured EDA, BVP, body temperature and 3-axis acceleration at sampling rates of 4, 64, 4, and 32 Hz, respectively. Participants also wore the Actiwave Cardio Monitor [29] on their chest, a miniature single channel ECG recorder with 512 Hz sampling rate. In the following discussions, the measures extracted based on the E4 and Actiwave signals will be referred to as wrist-worn physiological (WWP) and chest-worn physiological (CWP) measures, respectively. A Creative lavalier microphone was used to capture participants' speech during their presentations at 16 kHz sampling rate and 16-bit encoding. In total, this resulted in 10,800 minutes of acoustic and physiological data from 82 real and 216 VR presentations.

Prior to the presentation, participants completed several self-assessment reports which included the Trait-Scale of the Trait Anxiety Inventory (STAI) [30] and Communication Anxiety Inventory (CAI) [31], as well as the Personal Report of Public Speaking Anxiety (PRPSA) [32], capturing the general and communication-specific trait-based anxiety, the Big Five Inventory (BFI) [33], reflecting personality traits, the Brief Fear of Negative Evaluation (BFNE) [34], capturing feelings of apprehensions about others' evaluation, the Reticence Willingness to Communicate (RWTC) [35], reflecting reluctance to communicate, and two custom-made surveys of prior daily experiences (e.g., caffeine/alcohol/drug intake) and demographics (e.g., age, gender, ethnicity, degree, major). After each presentation, participants filled out another set of surveys including the State-Scale of CAI and STAI, and the State-Anxiety Enthusiasm (SAE), capturing state-based anxiety related to the preceding public speaking encounter, the Body Sensations Questionnaire (BSQ) [36], reflecting the body's response under stress, and a custom-made Presentation Preparation Performance (PPP) survey capturing degree of preparation and knowledge on the topic.

IV. Methodology

We will describe the data pre-processing (Section IV-A) and feature extraction (Section IV-B), and the various individual and contextual factors potentially contributing to PSA (Section IV-C). We will further outline the analyses carried out to answer the following three research questions: (1) Can we estimate PSA from bio-behavioral indices? (Section IV-D), (2) How do individual-specific factors contribute to PSA? (Section IV-E), and (3) Does systematic exposure through VR alleviate PSA? (Section IV-F).

A. Data Pre-processing

Outlier detection was performed for the EDA to detect potential dropouts. Outliers were defined as signal samples with values larger than three times the standard deviation from the median over an analysis window of 48 samples, a value visually yielding the best results. Outliers were replaced by carrying out a linear interpolation using the neighboring signal values. High-frequency noise was removed from the ECG signal using a low-pass finite impulse response filter of 45-samples length, followed by R-peak detection using the BioSPPy toolbox [37]. Voice activity detection (VAD) was performed on the audio signals to identify the presence and absence of speech using the Opensmile [38] toolbox.

B. Bio-behavioral Measures

A total of 7 WWP measures were extracted from the EDA, BVP, temperature, and acceleration signals. The Ledalab software [39] was used for EDA measures, including skin conductance level (SCL), skin conductance response (SCR) frequency, and mean amplitude of SCRs. Mean heart rate and mean inter-beat interval were measured from the BVP signal. Mean temperature and the $l2-$norm of the 3-axis acceleration signal were also computed. A total of 4 CWP features resulted from the ECG signal, including the root mean square of successive differences (RMSSD) of R-R intervals, the low-frequency (LF) and high-frequency (HF) energy of the ECG, and low-to-high frequency (LF:HF) ratio. The RMSSD and HF measures reflect the body's parasympathetic activity, which contributes to one's self-regulation ability [40]. While some studies have observed that LF and the LF:HF ratio are associated with sympathetic activity (i.e., fight-or-flight reaction), the role of these measures tends to be unclear [41]. Finally, 7 acoustic features were extracted from the audio data, which include the signal's root mean square energy, fundamental frequency, number of pauses, jitter, shimmer, zero crossing rate and voicing probability. These were computed over a 30-millisecond analysis window and were averaged over the speech segments of each audio file.

C. Moderating Factors

A total of 14 individual and contextual factors were used to model the inherently high variability across individuals and across various conditions, as obtained from the participants' self-reports (Section III). Significant differences between individuals with respect to their self-reports and bio-behavioral indices are studied based on these factors. These factors are further examined in terms of their ability to moderate the association between bio-behavioural indices and state-based PSA. Contextual factors include participants' self-reported level of preparation and knowledge on the presentation (PPP), as well as gender, age, native language, ethnicity, highest educational degree, degree currently being pursued, major, and recency of public speaking experience (Demographics). Individual factors include personality metrics (BFI) and trait-based general anxiety levels (STAI Trait).

D. Estimation of PSA from bio-behavioral indices

Linear regression was performed to estimate state-based PSA, as reported from the CAI State survey (Section III), based on each modality (CWP, WWP, Audio) and their combination (Section IV-B). Each regression model was evaluated through a leave-one-speaker-out (LOSO) cross-validation, according to which data from one speaker were included in the test set, while data from the remaining speakers were used for training. Estimated and the actual state-based PSA values were compared using Spearman's correlation.

E. Effect of individual-specific factors on PSA

1) Estimation of PSA from bio-behavioral measures augmented with individual and contextual factors: Linear regression was conducted based on the bio-behavioral features (CWP, WWP, Audio) and their combination with the individual and contextual factors. The original 18-dimensional feature vector of bio-behavioral indices was augmented by individual (trait-based anxiety from STAI, personality scores from BFI) and contextual factors (age, gender, native language, ethnicity, recency of public speaking experience, highest education achieved, currently pursuing degree). Each factor added one feature to the final feature set. The goal of each regression was to identify whether including individual indices affects the prediction of state-based anxiety based on the bio-behavioural measures captured from the different modalities. Each regression model was evaluated through a leave-one-speaker-out (LOSO) cross-validation by computing the Spearman's correlation between the actual and estimated state-based PSA.

2) Group-specific clustering: Different individuals are likely to experience different patterns of anxiety in various settings [5]. In order to integrate such individual and contextual differences into machine learning models, participants were clustered into different groups based on their individual and contextual factors (Section IV-C). This allows us to understand the subsets of factors affecting the state-based anxiety and its association with bio-behavioural signals. Principal Component Analysis (PCA) was applied on the above factors to reduce their dimensionality. Next, K-Means clustering was performed on the first two PCA dimensions, to obtain $K = 4$ groups of participants. The value of K was empirically determined based on the number of data samples.

3) Identifying PSA differences between groups of participants: Statistical analysis is used to identify significant differences between groups of participants with respect to their bio-behavioral indices and self-reported scores. Grouping is performed based on the individual and contextual indices

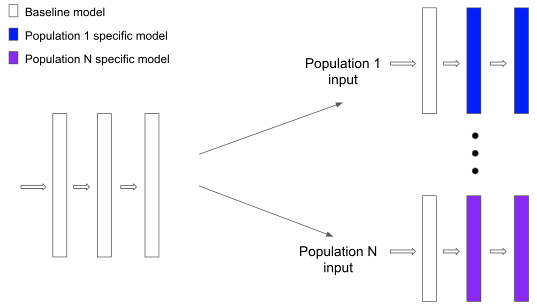


Fig. 1. Group-specific PSA models, implemented through feed-forward neural network (FNN) fine-tuning. A general FNN, trained on all participants, is adapted for each group, as defined by individual and contextual factors.

(Section IV-C). A t-test was used in the case of two groups, while an analysis of variance (ANOVA) was conducted when more than two groups were present.

4) Group-specific PSA models: A baseline feed-forward neural network (FNN) was trained based on all data samples to estimate state-based PSA from individuals' bio-behavioral indices. The FNN comprises of one hidden layer and is trained with a learning rate of 0.01, providing a general PSA estimation for all participant. Next, group-specific fine-tuning was performed, based on which samples from each group of participants were used to fine-tune the hidden and output layer of the baseline FNN, resulting in group-specific PSA estimations (Fig. 1). The learning rate during this fine-tuning was 0.001, providing fine-grain learning of the FNN parameters. Note that FNN fine-tuning is not performed for clusters with less than 3 data points, since this will not provide an adequate amount of data for re-training the last FNN layers. Three different types of group-specific PSA models were created based on three different group clustering criteria (individual, contextual, and their combination). Evaluation was performed through Spearman's correlation values using the state-based PSA estimations obtained from LOSO.

F. Comparing PSA before and after the VR sessions

We examine whether frequent exposure to VR stimuli, provided by the 8 VR sessions in TEST (Section III), affects participants' PSA. Statistical analysis through paired t-test was employed to determine significant differences between the PRE and the POST with respect to participants' trait-based and state-based self-reported anxiety, as well as their bio-behavioral measures (CWP, WWP, Audio).

V. Results

Results are reported separately for the PRE and POST treatments, since the TEST treatment (consisting of 8 VR sessions) that took place in between PRE and POST, renders them substantially different.

A. Estimation of PSA from bio-behavioral indices

Results indicate significant associations between the proposed bio-behavioural indices and the self-reported state-based anxiety scores (Table II). Combining physiological and acoustic features appears to be more useful compared to including measures from a single modality (Table II).

TABLE II
Spearman's correlation between actual and estimated state-based anxiety using wrist-worn physiological (WWP), chest-worn physiological (CWP), and acoustic measures

Bio-behavioural measures	**PRE session**	**POST session**
WWP	-0.05	-0.06
CWP	0.18	0.02
Acoustic	0.15	-0.37^{*}
WWP & CWP	0.32^{**}	0.25
WWP & Acoustic	0.14	-0.08
Acoustic & CWP	0.31^{**}	-0.19

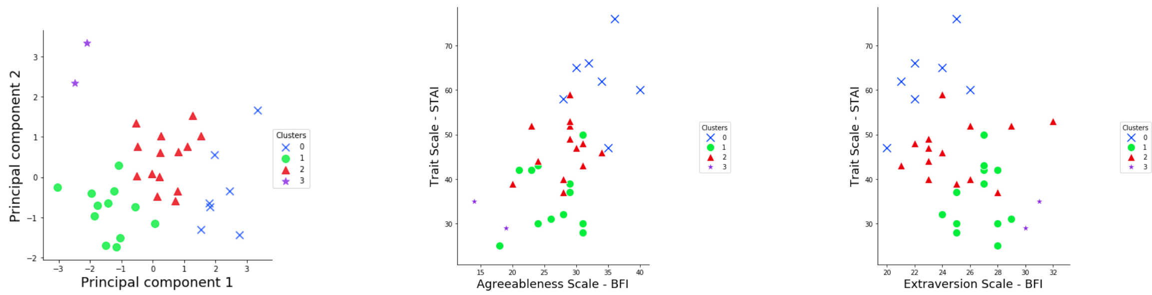
*: $p<0.05$, **: $p<0.01$

B. Effect of individual-specific factors on PSA

1) Estimation of PSA from bio-behavioral measures augmented with individual and contextual factors: Results from regression experiments indicate that augmenting the original bio-behavioral features with individual and contextual factors benefits the estimation of state-based PSA (Table IIIa). Individual factors related to general trait-based anxiety and personality when combined with CWP features, increase the accuracy of PSA estimation from 0.18 (Table II) to 0.36 ($p<0.01$) (Table IIIa) during the PRE. Similarly, augmenting the WWP and CWP features with individual factors increased Spearman's correlation from 0.32 ($p<0.01$) (Table II) to 0.38 ($p<0.01$) (Table IIIa). Significant increase was also found for the POST, benefiting most of the models which were previously relying solely on the bio-behavioral features. Similar benefits were provided by augmenting the bio-behavioral feature space with contextual factors (Table IIIb). Notably, combining acoustic features with information on age, gender, native language, degree currently pursued, and recency of public speaking increases Spearman's correlation from 0.15 (Table II) to 0.57 ($p<0.01$) (Table IIIb) in the PRE. Contextual factors did not benefit results on the POST, potentially due to the fact that the small number of data samples in the POST might undermine the robustness of our results.

2) Group-specific clustering: K-Means clustering suggests the presence of various groups of participants. Fig. 2a depicts four separable clusters based on all individual factors. Visual inspection of the resulting clusters indicates groups of participants with high trait anxiety and high agreeableness (Fig. 2a), as well as high trait anxiety and low extraversion (Fig. 2b). Clustering based on contextual factors provided similar plots.

3) Identifying PSA differences between groups of participants: Significant differences can be observed among participant groups based on the various individual and contextual factors with respect to their PSA (Table IV). Participants who had given a presentation 4-8 times in the last 3 months, reported significantly higher PPP scores (mean=17.80, stand. dev=3.35) compared to participants who had presented only 1-3 times during the same duration (mean=12.85, std=3.90). Undergraduate students reported significantly higher trait anxiety (CAI trait) (mean=15.33, stand. dev=3.53) and depict higher SCR frequency (mean=13.02, stand. dev=3.77) compared to graduate students (CAI trait; mean=10.80, stand. dev=2.28) (EDA frequency; mean=6.10, stand. dev=4.23). Participants of Asian ethnicity depicted increased shimmer in their speech



(a) PC1 vs PC2 (b) Trait anxiety vs Agreeableness (c) Trait anxiety vs Extraversion

Fig. 2. (a) Groups of participants as determined by K-Means clustering performed on the first two principal component dimensions of the individual factors. (b)-(c) Pairwise plots of individual factors: trait anxiety and Big Five Inventory (BFI) personality metric agreeableness/extraversion with grouping governed by K-means clustering on principal component dimensions of all the individual factors

TABLE III
SPEARMAN'S CORRELATION BETWEEN ACTUAL AND PREDICTED STATE-BASED ANXIETY BASED ON LINEAR REGRESSION PER MODALITY AND THEIR COMBINATION WITH INDIVIDUAL/CONTEXTUAL FACTORS

(a) Bio-behavioral measures augmented with individual factors related to trait-based (T) and personality (P) scores

Bio-behavioural measures	Individual factors	PRE session	POST session
WWP	T,P	-0.77^{**}	0.54^{**}
CWP	T,P	0.36^{**}	0.60^{**}
Acoustic	T,P	0.11	0.51^{**}
WWP & CWP	T,P	0.38^{**}	0.50^{**}
WWP & Acoustic	T,P	0.01	0.35
Acoustic & CWP	T,P	0.22	0.38^{*}

(b) Bio-behavioral measures augmented with contextual factors related to age (A), gender (G), native language (L), ethnicity (E), recency of public speaking presentation (R), highest degree earned (H), and degree currently pursuing (D)

Bio-behavioural measures	Contextual factors	PRE session	POST session
WWP	H	0.47^{**}	-0.20
CWP	A,R	0.49^{**}	0.24
Acoustic	A,G,L,D,R	0.57^{**}	-0.35
WWP & CWP	A,H,R	0.36^{*}	-0.22
WWP & Acoustic	L,E,H,R	0.36^{*}	-0.54^{**}
Acoustic & CWP	L,E,A,R	0.62^{**}	-0.33

*: $p<0.05$, **: $p<0.01$

(mean=0.13, stand. dev=0.03) compared to White/Caucasian participants (mean=0.12, stand. dev=0.02), potentially due to general phonological differences between the two groups. Hispanic/Latino participants reported significantly higher BSQ (mean=2.15, stand. dev=0.70) compared to White/Caucasian participants (mean=1.78, stand. dev=0.55), as well as a higher SAE (mean=60.77, stand. dev=8.92) compared to African American participants (mean=47.83, stand. dev=7.25). Male participants depicted higher shimmer (mean=0.14, stand. dev=0.03) compared to females (mean=0.12, stand. dev=0.02).

4) Group-specific PSA models: Results obtained from the group-specific FNNs provide better performance compared to the general FNN (Table V). While FNNs refined based on individual-specific clustering marginally improve the Spearman's correlation, FNNs refined using individual and contextual-specific clusters depict significant benefits, yielding a final Spearman's correlation of 0.55 ($p<0.05$) compared to 0.10 from the general FNNs. It is important to note that the sample size for the FNN decreased when combining the individual and contextual factors, because of missing data for some participants. This imbalance was taken into account and the context-based FNN was also tested on the reduced data set, which included the 18 participants whose individual and contextual metrics were both available, and results were found to be consistent with Table V.

C. Comparing PSA before and after the VR sessions

Significant differences with respect to self-reported and bio-behavioral indices were found before and after the 8 VR sessions (TEST treatment). The corresponding measures were obtained during public speaking presentations in front of a real audience, which occurred before (PRE treatment) and after (POST treatment) the VR sessions. Results suggest a significant reduction in terms of self-reported state-based PSA (CAI, SAE) between the PRE and the POST, indicating that participants felt less stressed when presenting in front of the real audience after experiencing the VR sessions. It is also noteworthy that participants reported a reduction with approaching significance ($p = 0.06$) in terms of trait-based PSA, as obtained from the PRPSA metric, which might suggest a long-term usage of the proposed VR exposure. Our results further reflect a significant reduction in SCR frequency and heart rate between the PRE and the POST, suggesting a reduction in the amount of sympathetic activity related to the fight-or-flight response during the POST. Although there are significant differences between the PRE and POST treatments with respect to the jitter and shimmer, the difference is not in the expected direction. Jitter, a measure related to the variations of fundamental frequency and speech breathiness [42], has increased during the POST compared to the PRE. This might be due to the fact that participants might have been more eager to touch upon as many discussion points as possible and show improved public speaking skills in front of the real audience during the POST, which might have caused the increased breathiness in their voice. While the limited number of samples ($n=27$) based on which this analysis is performed does not provide conclusive results, these findings indicate that systematic exposure to the public speaking through VR stimuli might be able to alleviate PSA.

TABLE IV
ANOVA & T-TEST RESULTS FOR MEASURING SIGNIFICANT DIFFERENCES IN PUBLIC SPEAKING ANXIETY BETWEEN VARIOUS GROUPS OF INDIVIDUALS WITH RESPECT TO SELF-REPORTS AND BIO-BEHAVIOURAL INDICES

a) ANOVA and T-test results for self-assessments

	Communication Anxiety Inventory (CAI) trait (Dyadic)	Brief fear of Negative Evaluation (BFNE)	Retience Wilingness to Communicate (RWTC)	Communication Anxiety Inventory (CAI) state	Body sensations questionnaire (BSQ)	Post Presentation Performance (PPP)	State Anxiety Enthusiasm (SAE)
Age	$f(2,50)=0.80$	$f(2,50)=0.07$	$f(2,50)=0.34$	$f(2,50)=1.57$	$f(2,50)=1.31$	$f(2,50)=0.18$	$f(2,50)=1.77$
Gender	$t(50)=0.49$	$t(50)=-1.23$	$t(50)=-1.61$	$t(50)=-0.42$	$t(50)=-0.57$	$t(50)=0.24$	$t(50)=-1.17$
Ethnicity	$f(4,50)=0.51$	$f(4,50)=0.87$	$f(4,50)=0.57$	$f(4,50)=2.11$	$f(4,50)=\mathbf{3.49}^{**}$	$f(4,50)=0.61$	$f(4,50)=\mathbf{3.27}^{*}$
College	$f(8,50)=0.35$	$f(8,50)=\mathbf{2.16}^{*}$	$f(8,50)=1.09$	$f(8,50)=1.25$	$f(8,50)=0.35$	$f(8,50)=0.57$	$f(8,50)=1.01$
Native language	$f(3,50)=1.85$	$f(3,50)=0.42$	$f(3,50)=0.85$	$f(3,50)=0.26$	$f(3,50)=0.34$	$f(3,50)=2.07$	$f(3,50)=0.22$
Highest education	$f(3,50)=\mathbf{2.95}^{*}$	$f(3,50)=0.19$	$f(3,50)=0.14$	$f(3,50)=0.07$	$f(3,50)=0.17$	$f(3,50)=0.44$	$f(3,50)=0.15$
Presentation in last 3 months	$f(3,50)=1.35$	$f(3,50)=2.04$	$f(3,50)=0.99$	$f(3,50)=1.91$	$f(3,50)=1.18$	$f(3,50)=\mathbf{8.04}^{**}$	$f(3,50)=\mathbf{4.22}^{**}$

* $p<0.05$. ** $p<0.01$

b) ANOVA and T-test results for bio-behavioural indices

	Body temperature	Skin conductance response (SCR) frequency	Root mean square of successive differences (RMSSD) of R-R intervals	Speech jitter	Speech shimmer
Age	$f(2,50)=1.72$	$f(2,50)=1.77$	$f(2,50)=0.44$	$f(2,50)=0.16$	$f(2,26)=0.26$
Gender	$t(50)=-0.048$	$t(50)=1.18$	$t(50)=0.97$	$t(50)=1.45$	$t(50)=\mathbf{2.20}^{*}$
Ethnicity	$f(4,50)=\mathbf{2.66}^{*}$	$f(4,50)=1.71$	$f(4,50)=1.08$	$f(4,50)=\mathbf{3.11}^{*}$	$f(4,50)=0.88$
Native language	$f(3,50)=2.15$	$f(3,50)=\mathbf{3.05}^{*}$	$f(3,50)=0.08$	$f(3,50)=0.69$	$f(3,50)=1.12$
Highest education	$f(3,50)=1.51$	$f(3,50)=\mathbf{4.74}^{**}$	$f(3,50)=2.06$	$f(3,50)=0.53$	$f(2,26)=0.32$
College	$f(8,50)=1.25$	$f(8,50)=1.41$	$f(8,50)=0.40$	$f(8,50)=0.49$	$f(8,50)=0.92$
Presentation in last 3 months	$f(3,50)=1.32$	$f(3,50)=0.43$	$f(3,50)=\mathbf{7.57}^{**}$	$f(3,50)=0.25$	$f(3,50)=0.47$

* $p<0.05$. ** $p<0.01$

TABLE V
SPEARMAN'S CORRELATION BETWEEN THE ACTUAL AND PREDICTED STATE-BASED ANXIETY MEASURES BASED ON THE GROUP-SPECIFIC FEED-FORWARD NEURAL NETWORK (FNN) MODELS

Type of FNN model	Sample size	PRE session
Context-based FNN	35	-0.31
Baseline FNN	35	-0.39^{*}
individual FNN	21	0.15
Baseline FNN	21	0.02
Context & Trait based FNN	18	0.55^{*}
Baseline FNN	18	0.10

*: $p<0.05$, **: $p<0.01$

TABLE VI
T-TEST RESULTS COMPARING SIGNIFICANT DIFFERENCES BETWEEN PRE AND POST, BEFORE AND AFTER THE VR SESSIONS, WITH RESPECT TO SELF-REPORTED AND BIO-BEHAVIORAL MEASURES

Self-reported measures	**PRE session**	**POST session**	**T-test results**
Communication Anxiety Inventory (CAI), State Scale	46.25	39.74	$t(26)=2.33^{**}$
State-Anxiety Enthusiasm Scale (SAE)	55.66	48.14	$t(26)=2.69^{**}$
Personal Report of Public Speaking Anxiety (PRPSA)	104.85	92.00	$t(26)=1.88^{\dagger}$
Bio-behavioral measures	**PRE session**	**POST session**	**T-test results**
Skin conductance response frequency	11.83	6.84	$t(26)=3.33^{**}$
Heart rate	89.23	82.46	$t(26)=2.28^{*}$
Body temperature	32.71	31.80	$t(26)=1.85^{\dagger}$
Jitter	0.02	0.04	$t(26)=-2.84^{**}$
Shimmer	0.12	0.15	$t(26)=-2.96^{**}$

†: $p<0.1$, *: $p<0.05$, **: $p<0.01$

VI. DISCUSSION

The results reported in this paper should be considered in light of the following limitations. Current analysis is based solely on the data obtained during PRE and the POST treatments, investigating the data from VR sessions might shed more light on our findings. As part of our future work, we plan to examine whether self-reported and bio-behavioral measures are affected by the various VR settings (e.g., audience size and reactions, room conditions), and track their progress across the 8 VR sessions. In addition, our analysis so far has not examined participants' visual cues, such as facial expressions or body gestures, an important channel reflecting the degree of PSA. Finally, we have not yet accounted for the fact that the differences found between the PRE and the POST treatments might be attributed to the habituation arising from conducting the 10 public speaking tasks in a relatively small span of time (i.e., 2 weeks). A possible future direction of this research will include comparing desensitization through VR stimuli with other forms of interventions (e.g., desensitization with real audience, desensitization combined with cognitive restructuring feedback) in order to understand whether such differences will still be present. Cognitive restructuring feedback is a method which aims to modify an individual's negative perception of a threatening stimuli [43], [44]. Wearable and mobile devices can now afford us a unique solution to provide cognitive restructuring feedback in-the-moment, when it is needed the most. The hypothesis is that such an in-the-moment feedback would be able to change an individual's thought process on the fly, suppress their irrational fears, and direct them toward a healthier perception of public speaking. In our future work, we aim to design a system where bio-behavioural indices and the group-specific PSA models are used to predict individuals' state-based PSA in real-time and provide them with in-the-moment feedback. The work presented in this paper lays the foundation to work toward this future direction by understanding the different bio-behavioral expressions of PSA among individuals and designing group-specific machine learning models capable of taking these factors into account.

VII. CONCLUSION

This paper examined quantifiable estimators of PSA and the effect of VR in alleviating anxiety during public speaking. Statistical analysis indicates high inter-individual variability in the way participants perceive and experience PSA. Incorporating individual and contextual factors into machine learning models–either in the feature space, or through model adaptation–can improve PSA estimation. Results demonstrate that systematic exposure to public speaking, implemented via VR, can help alleviate PSA in terms of self reports and bio-behavioral indices. In our future work, we will analyze the data from the presentation sessions in the VR environment. Along with the individual-specific factors we will also integrate cognitive aptitude in the PSA models. Finally, we will obtain momentary PSA annotations from observational coding and design systems that can predict PSA in real-time, which will provide the foundation for in-the-moment PSA interventions.

References

[1] N. E. Betz, "Contributions of self-efficacy theory to career counseling: A personal perspective," *The Career Development Quarterly*, vol. 52, no. 4, pp. 340–353, 2004.

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

[3] N. C. C. for Mental Health *et al.*, *Social anxiety disorder: the NICE guideline on recognition, assessment and treatment*. Royal College of Psychiatrists, 2013.

[4] K. Boonkit, "Enhancing the development of speaking skills for non-native speakers of english," *Procedia-social and behavioral sciences*, vol. 2, no. 2, pp. 1305–1309, 2010.

[5] 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.

[6] 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*. Springer, 2013, pp. 116–128.

[7] 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 *Proc. ICMI*. ACM, 2014, pp. 200–203.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] M. M. North, S. M. North, and J. R. Coble, "Virtual reality therapy: An effective treatment for psychological disorders." 2002.

[20] ——, "Virtual reality therapy: An effective treatment for psychological," *Virtual reality in neuro-psycho-physiology: Cognitive, clinical and methodological issues in assessment and rehabilitation*, vol. 44, p. 59, 1997.

[21] D. Pertaub, M. Slater, and C. Barker, "An experiment on fear of public speaking in virtual reality," *Studies in health technology and informatics*, pp. 372–378, 2001.

[22] R. R. Behnke and C. R. Sawyer, "Anticipatory anxiety patterns for male and female public speakers," *Communication education*, vol. 49, no. 2, pp. 187–195, 2000.

[23] 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.

[24] 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.

[25] 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*. ACM, 2015, pp. 1143–1154.

[26] M. Chollet, T. Wörtwein, L.-P. Morency, and S. Scherer, "A multimodal corpus for the assessment of public speaking ability and anxiety." in *LREC*, 2016.

[27] "Presentation simulator," http://www.presentationsimulator.com/, Accessed 1/20/2019.

[28] "Empatica e4," https://www.empatica.com/, Accessed 1/20/2019.

[29] "Actiwave cardio monitor," https://www.camntech.com/, Accessed 1/20/2019.

[30] C. D. Spielberger, "State-trait anxiety inventory," *The Corsini encyclopedia of psychology*, pp. 1–1, 2010.

[31] 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.

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

[33] O. P. John, S. Srivastava *et al.*, "The big five trait taxonomy: History, measurement, and theoretical perspectives," *Handbook of personality: Theory and research*, vol. 2, no. 1999, pp. 102–138, 1999.

[34] 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, p. 42, 2009.

[35] 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.

[36] 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.

[37] "Biosppy bio signal processing in python," https://biosppy.readthedocs.io/en/stable/, Accessed 1/20/2019.

[38] "Opensmile audeering," https://www.audeering.com/opensmile/, Accessed 1/20/2019.

[39] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *Journal of neuroscience methods*, vol. 190, no. 1, pp. 80–91, 2010.

[40] S. Laborde, E. Mosley, and J. F. Thayer, "Heart rate variability and cardiac vagal tone in psychophysiological research–recommendations for experiment planning, data analysis, and data reporting," *Frontiers in psychology*, vol. 8, p. 213, 2017.

[41] G. E. Billman, "The lf/hf ratio does not accurately measure cardiac sympatho-vagal balance," *Frontiers in physiology*, vol. 4, p. 26, 2013.

[42] J. Hillenbrand and R. A. Houde, "Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech," *Journal of Speech, Language, and Hearing Research*, vol. 39, no. 2, pp. 311–321, 1996.

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

[44] 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.