

Knowledge- and Data-Driven Models of Multimodal Trajectories of Public Speaking Anxiety in Real and Virtual Settings

Ehsanul Haque Nirjhar
Texas A&M University
College Station, Texas, USA
nirjhar71@tamu.edu

Amir H. Behzadan
Texas A&M University
College Station, Texas, USA
abehzadan@tamu.edu

Theodora Chaspari
Texas A&M University
College Station, Texas, USA
chaspari@tamu.edu

ABSTRACT

Public speaking skills are essential to professional success. Yet, public speaking anxiety (PSA) is considered one of the most common social phobias. Understanding PSA can help communication experts identify effective ways to treat this communication-based disorder. Existing works on PSA rely on self-reports and aggregate multimodal measures which do not capture the temporal variation in PSA. This paper examines temporal trajectories of acoustic and physiological measures throughout the public speaking encounter with real and virtual audiences, and aims to model those in both knowledge- and data-driven ways. Knowledge-driven models leverage theoretically-grounded patterns through fitting interpretable parametric functions to the corresponding signals. Data-driven models consider the functional nature of multimodal signals via functional principal component analysis. Results indicate that the parameters of the proposed models can successfully estimate individuals' trait anxiety in both real-life and virtual reality settings, and suggest that models trained on data obtained in virtual public speaking stimuli are able to estimate levels of PSA in real-life.

CCS CONCEPTS

• **Human-centered computing** → Ubiquitous and mobile computing; • **Computing methodologies** → Supervised learning.

KEYWORDS

Public speaking anxiety, physiological signals, speech, functional principal component analysis, time trajectory

ACM Reference Format:

Ehsanul Haque Nirjhar, Amir H. Behzadan, and Theodora Chaspari. 2021. Knowledge- and Data-Driven Models of Multimodal Trajectories of Public Speaking Anxiety in Real and Virtual Settings. In *Proceedings of the 2021 International Conference on Multimodal Interaction (ICMI '21)*, October 18–22, 2021, Montréal, QC, Canada. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3462244.3479964>

1 INTRODUCTION

Skillful public speaking is key to personal and professional success, and a necessary component of social interaction, since it allows individuals to demonstrate their ideas and share their knowledge [25]. However, many people do not feel comfortable speaking

in front of an audience [13]. This fear of public speaking is termed as “public speaking anxiety” (PSA). PSA is a communication-based phobia [29], often eliciting negative thoughts and feelings of apprehension, while causing increased heart rate and sweat [6, 7]. Quantifying temporal trajectories of PSA throughout the public speaking encounter can provide valuable insights into the nature of the disorder.

Various studies in psychological and cognitive sciences have explored PSA [1, 2, 5, 28]. These studies tend to elicit PSA among participants using real-life settings (e.g. no audience, small sized audience) [8, 25] or virtual reality (VR) settings (e.g. virtual audience projection, 2D avatar) [4, 14, 23]. The majority of prior studies rely on conventional self-reported indices [17, 27], which do not capture momentary variations in PSA. Recent works have also explored multimodal indices (e.g. heart rate, acoustic properties of speech, body gestures) for modeling PSA [9, 19, 30]. Despite the valuable findings on PSA, existing work often uses aggregated multimodal measures over the entire public speaking session, which might cause the loss of valuable temporal information [4, 8]. PSA has been theoretically postulated to depict temporal variations throughout the public speaking task. Temporal variations include decreasing (i.e., habituation) or increasing (i.e., escalation) patterns over time, as well as an increase followed by a decrease (i.e., sensitization) [6]. These trajectories have been found to differ across individuals. For example, individuals with high trait anxiety and high sensitivity to public speaking depict stronger escalation or sensitization patterns compared to their counterparts [7]. Few studies have addressed this issue by developing computational models of PSA from the knowledge base acquired in psychology and communication [19, 26]. Detailed exploration of multimodal trajectories over time can potentially help understand causes and antecedents of PSA. Moreover, prior works primarily model PSA either in real-life [8, 19] or VR settings [4, 14]. There has not been much work exploring the extent to which multimodal indices of PSA from one setting (e.g., VR) can be used to model PSA in another setting (e.g., real-life). Since obtaining data in VR is convenient and cost-effective, it is important to understand whether time trajectories of PSA in VR can be used to model PSA in real-life.

We model multimodal temporal trajectories of speech and physiology from data obtained in real-life and VR public speaking settings through knowledge- and data-driven approaches. Knowledge-driven trajectories build upon previously identified patterns of escalation, sensitization, and habituation during public speaking [6], and are modeled through parametric temporal functions [19]. Data-driven trajectories exploit the functional nature of multimodal signals by identifying common dominant temporal dimensions of variation, and are obtained through functional principal component

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).

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

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8481-0/21/10...\$15.00

<https://doi.org/10.1145/3462244.3479964>

analysis (fPCA) [18, 31]. In this work, we examine the knowledge- and data-driven models in association with individuals' trait anxiety to identify ways in which individuals of different psychological characteristics experience PSA. Results evaluated through 71 real-life and 232 VR public speaking sessions indicate that measures obtained by the data- and knowledge-driven models, and their combination, can reliably estimate trait anxiety with moderate Spearman's correlation values (i.e., $\rho = 0.48$, $p < 0.01$) between the actual and estimated trait anxiety scores. Results further suggest that PSA models learnt in VR settings can be applied to real-life settings. Implications from this work can help toward better understanding sources of PSA and improving personalized training for promoting a healthy perception of public speaking.

2 RELATED WORKS

The majority of prior work views PSA as a state anxiety, momentarily caused by specific stimuli [5]. Chen *et al.* [8] employed a data-driven model using a multimodal feature set to predict audience evaluation scores from acoustic measures. Batrinca *et al.* [4] employed aggregate multimodal indices to estimate state PSA. Yadav *et al.* [29, 30] incorporated contextual factors with multimodal indices to predict state anxiety. Using these contextual factors, their work further proposed group-based prediction models to estimate PSA. Beyond aggregate measures, Bodie [6] introduced the notion of salient temporal patterns of PSA (e.g. habituation, sensitization). He further discussed the importance of trait-based anxiety, referring to one's general tendency to experience anxiety, and explained how trait anxiety moderates temporal trajectories of state-based momentary PSA responses. Inspired by this, Nirjhar *et al.* proposed computational models to represent various trajectories of physiological and acoustic features and demonstrated that these trajectories can be related to trait-based anxiety [19]. While these approaches build computational knowledge-driven models, data-driven approaches can facilitate the process by extracting salient modes of variation from a large dataset. Metallinou *et al.* [18] used fPCA to quantify atypicality in facial expressions for children with autism, while Arias *et al.* [3] detected deviation of emotional from neutral speech with fPCA. In this work, we augment a fPCA-based data-driven models with knowledge-driven models proposed in [19] and examine how they can be used to effectively estimate PSA in real-life and VR settings.

This paper advances the current state of knowledge in the following ways: (1) Rather than aggregate measures, we study knowledge- and data-driven models of multimodal signal trajectories (i.e., physiology, acoustic measures) and their association with PSA; (2) We examine how temporal patterns of PSA manifested through multimodal trajectories can estimate the trait anxiety in different settings (i.e. real-life, VR); and (3) We investigate the extent to which multimodal time trajectories learned by VR data can be transferred to real-life.

3 DATA

3.1 Dataset Description

For our experiment, we used the publicly available VerBIO dataset [30], which contains audio recordings, physiological signals, and self-reported measures from 344 public speaking sessions completed by 55 participants. These participants performed public speaking tasks in front of real-life and VR audiences. They completed 10 public

speaking sessions (i.e., 2 in real-life, 8 in VR) over four different days. In Days 1 and 4, participants performed a single session in front of a real-life audience. Days 2 and 3 each involved 4 sessions of public speaking in front of a VR audience, where participants wore an Oculus Rift headset [24] to be immersed in the VR environment generated by Virtual Orator software [20]. During each session, participants delivered an oral presentation (mean duration 4.5 minutes) about a news article from a randomly assigned topic (e.g. technology, entertainment). Participants wore the Empatica E4 wristband [10] and Creative lavalier microphone, which captured physiological and speech signals, respectively. The E4 wristband recorded electrodermal activity (EDA) and average heart rate (HR), sampled at 4 and 1 Hz, respectively. The microphone acquired the audio signal at 16 kHz sampling rate and 16-bit encoding. Participants completed a number of standardized surveys at each day of the user study, including the Brief Fear of Negative Evaluation (BFNE) [27] and Personal Report of Public Speaking Anxiety (PRPSA) [17]. BFNE records the discernment caused by worrying about audience evaluation, while PRPSA measures the communication-specific trait-anxiety. These measures serve as an estimate of participants' trait-based anxiety. The ranges of BFNE and PRPSA scores are 12–60 and 34–170, respectively, where higher score refers to higher anxiety. The mean BFNE and PRPSA in our data are 38.6 (SD=9.8) and 95.8 (SD=24.3), respectively. We excluded the sessions which do not contain all the bio-behavioral indices and self-reported scores due to technical errors. Finally, data from a total of 71 real-life and 232 VR sessions are used for further experimentation in this paper. We assume that each session can be considered as an independent data sample, as participants spoke about different topics in front of different audience (or VR environment) in different sessions.

3.2 Multimodal Feature Time series

Noise suppression and outlier removal were performed on the physiological signals, and voice activity detection (VAD) was performed on the audio [29]. Several multimodal features are extracted from these signals. Physiological features include the mean skin conductance level (SCL) and average heart rate (HR), while acoustic features contain fundamental frequency (F0), jitter, and shimmer. Acoustic features are extracted using OpenSMILE [11]. These features have been extensively used in prior works related to stress and anxiety [4, 15, 16]. These measures were computed over 15-second non-overlapping time intervals, which constitute the multimodal feature time series. We performed min-max normalization on physiological and acoustic features for each public speaking session.

4 METHODOLOGY

We describe the modeling of the multimodal time series using knowledge-driven (Section 4.1) and data-driven trajectories (Section 4.2). Next, we present how these trajectory parameters were used to estimate PSA in real-life and VR. Finally, we describe the prediction of PSA in real-life using models trained on data from the VR setting (Section 4.3).

4.1 Knowledge-driven trajectories

Previous research has identified distinct multimodal patterns associated with PSA [6, 19]. Following theoretically postulated habituation and sensitization trajectories by [6], we use a quadratic polynomial model and a transient oscillation model to represent

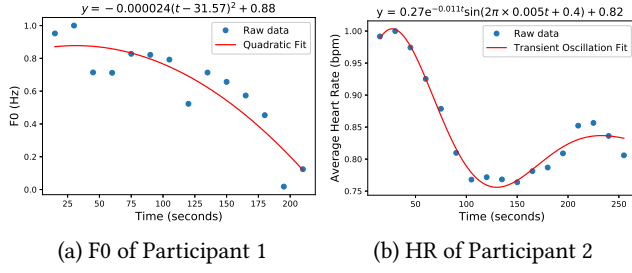


Figure 1: Knowledge-driven trajectories of fundamental frequency (F0) and heart rate (HR) for two participants. Model parameters are presented in the equation above each plot.

the physiological and acoustic time series obtained in each public speaking session (Section 3.2). Visual inspection of the F0, jitter, shimmer, and SCL measures indicated that the corresponding time series exhibit quadratic polynomial patterns (Fig. 1(a)), therefore these were fitted into (1):

$$y(t) = a(t - t_o)^2 + b \quad (1)$$

where $y(t)$ is the feature value at time t , a refers to the rate of temporal change, t_o is the time when minima (or maxima) occurs, and b is the bias term. Positive values of a are indicative of habituation, while negative a indicates sensitization effects.

Meanwhile, HR time series depicted oscillatory patterns (Fig. 1(b)), and therefore were fitted into the transient oscillation model, represented by (2):

$$y(t) = Ae^{-\tau t} \sin(2\pi ft + \phi) + c \quad (2)$$

where A , τ , f , ϕ , and c correspond to amplitude, decay rate, oscillation frequency, initial phase, and bias. Small values of τ indicate a slower dampening of the oscillation, which resembles slow rate of habituation.

The quadratic polynomial model has three parameters, while the transient oscillation model has five. Therefore, acoustic and physiological trajectories have nine and eight parameters, respectively, for each public speaking session.

4.2 Data-driven trajectories

In order to capture salient patterns of the multimodal feature time series in a data-driven approach, we employ fPCA [31], a time series version of PCA that determines the dominant components of variation from a set of time series by maximizing the projected variance. According to fPCA, the i^{th} time series $y_i(t)$ of a dataset is decomposed into a set of Gaussian basis functions $\phi_1(t), \phi_2(t), \dots, \phi_K(t)$:

$$y_i(t) = \sum_{k=1}^K C_{ik} \phi_k(t) + \varepsilon \quad (3)$$

where K is the total number of components, C_{ik} is the fPCA score of the k^{th} component of i^{th} time series, and ε is the reconstruction error. We learned the fPCA components using the time series of each multimodal measure from real-life and VR settings using open-source PACE library [21]. K is selected on the basis of fraction of explained variance. For each feature time series, the first K fPCA components that captured 99% variability in the data were retained. The fPCA scores C_{ik} along with the root-mean-square error (i.e., serving as a measure of confidence of the estimated parameters) were calculated for each multimodal feature. This results in a data-driven trajectory feature vector of dimensionality 12 and 9 for the acoustic and physiological features, respectively. Fig. 2 shows 1^{st}

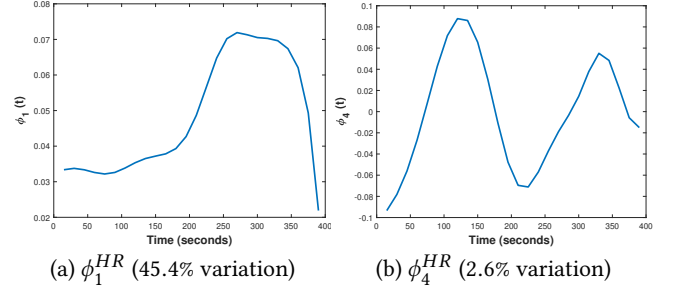


Figure 2: 1^{st} and 4^{th} basis functions obtained from heart rate (HR) time series via functional principal component analysis (fPCA).

and 4^{th} basis functions of HR time series. It is noteworthy that these functions mimic sensitization and oscillatory patterns.

4.3 Estimating trait anxiety from trajectories

Using combinations of the parameters from the knowledge- and data-driven trajectory representations of physiological and acoustic features, we perform regression experiments to predict PSA in real-life (Experiment 1) and VR (Experiment 2), separately. In addition, we estimate PSA in real-life by training a regression model on data from the VR setting (Experiment 3). In all experiments, random forest (RF) regression (RandomForestRegressor function from scikit-learn [22]) is used to estimate trait anxiety (i.e., BFNE, PRPSA; Section 3.1) using the aforementioned multimodal knowledge- and data-driven temporal parameters. Hyper-parameter tuning for all experiments was performed in terms of optimal number of trees and maximum tree depth. This resulted in 50 trees and 25 maximum tree depth of the random forest. For Experiment 3, we used these best performing hyperparameters yielding from the VR setting to predict PSA in real-life. For evaluating our approach, we employed a 10-fold cross-validation framework, repeated 50 times to counter potential randomness from the RF. Given the ordinal nature of the outcome, the Spearman's correlation between the actual and estimated trait scores is used as the evaluation metric.

5 RESULTS

5.1 Experiments 1-2: Estimating PSA in real-life and VR

Results obtained in Experiments 1 and 2 for estimating trait anxiety using features from the knowledge- and data-driven trajectory models are summarized in Tables 1a and 1b. Knowledge-driven trajectory parameters depict moderate performance for estimating BFNE in real-life when physiological ($\rho = 0.25, p < 0.05$) and both ($\rho = 0.18, p = 0.14$) modalities are used. These results show improvement in the case where knowledge- and data-driven parameters are combined for both modalities ($\rho = 0.35, p < 0.01$; Table 1a). In real-life settings, data-driven parameters alone do not appear to be good predictors of trait anxiety. However, these parameters perform significantly better in predicting BFNE and PRPSA in the VR setting. The acoustic modality alone ($\rho = 0.37, p < 0.01$) and in combination with the physiological modality ($\rho = 0.48, p < 0.01$) depict better performance in predicting BFNE in VR. Similar results are obtained for predicting PRPSA with both modalities (Table 1b). This performance discrepancy between VR and real-life settings might be potentially attributed to the large number of samples in

Table 1: Spearman’s correlation coefficient between actual and predicted trait anxiety scores using various modalities.**(a) Experiment 1: Estimating PSA in real-life setting**

Modality	Trait	Knowledge-driven	Data-driven	Both
Physiological	BFNE	0.25	0.09	0.25
	PRPSA	-0.20	-0.05	-0.13
Acoustic	BFNE	-0.06	0.25*	0.19
	PRPSA	-0.08	-0.13	-0.15
Both	BFNE	0.18	0.18	0.35**
	PRPSA	-0.24*	-0.14	-0.25

*: $p < 0.05$, **: $p < 0.01$ **(b) Experiment 2: Estimating PSA in virtual reality (VR) setting**

Modality	Trait	Knowledge-driven	Data-driven	Both
Physiological	BFNE	0.27**	0.27**	0.30**
	PRPSA	0.20*	0.25**	0.25**
Acoustic	BFNE	0.10	0.37**	0.30**
	PRPSA	0.02	0.28**	0.22**
Both	BFNE	0.27**	0.48**	0.46**
	PRPSA	0.20**	0.39**	0.34**

*: $p < 0.05$, **: $p < 0.01$ **(c) Experiment 3: Estimating PSA in real-life using VR model**

Modality	Trait	Knowledge-driven	Data-driven	Both
Physiological	BFNE	-0.09	0.09	0.04
	PRPSA	0.17	0.24	0.26
Acoustic	BFNE	0.07	-0.03	0.01
	PRPSA	0.07	0.27*	0.27
Both	BFNE	-0.11	-0.01	-0.04
	PRPSA	0.07	0.26*	0.20

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

the first ($N = 232$) compared to the second ($N = 71$). The combination of knowledge- and data-driven trajectory parameters with both modalities can effectively predict BFNE ($\rho = 0.46$, $p < 0.01$) and PRPSA ($\rho = 0.34$, $p < 0.01$). These results indicate that the physiological modality performs well with knowledge-driven parameters, while the acoustic modality favors the data-driven parameters. Combining both types of trajectory parameters can further increase the predictive power of the model. Physiological features (i.e. SCL, HR) inherently depict structured patterns, which can be captured through the knowledge-driven parameters. On the other side, acoustic features are more diverse, therefore it is likely that data-driven parameters are useful for modeling those measures. Finally, we anticipate that the negative correlations in estimating PRPSA (Table 1a) are due to the fact that PRPSA captures a wider range of public speaking anxiety symptoms than BFNE, therefore it is difficult to quantify with the small number of real-life samples.

5.2 Experiment 3: Estimating PSA in real-life using model trained on the VR data

Results from Experiment 3, shown in Table 1c indicate that although BFNE in real-life is very effectively estimated using the RF models trained on the real-life data, the corresponding outcome is not accurately estimated using the RF model trained on the VR data. On the other hand, PRPSA in real-life is reliably estimated using the VR model, especially in the case of the physiological modality with the data-driven ($\rho = 0.27$, $p < 0.05$) and both ($\rho = 0.26$, $p < 0.05$) trajectory parameters. For all combinations, predicting PRPSA in real-life using data from VR outperforms the models trained on the real-life data (Table 1a). This suggests that the model trained using data from VR sessions can be useful in predicting PSA in naturalistic scenarios. A potential reason explaining the discrepancy in performance between BFNE and PRPSA lies in the fact that the first is highly dependent on the context (i.e., audience, room), which

changes between the VR and real-life. However, PRPSA captures general trait anxiety in communication settings, therefore it might reflect psychological traits independently from contextual factors.

6 DISCUSSION

Results from this paper indicate that acoustic and physiological trajectories obtained from public speaking tasks are indicative of individuals’ trait anxiety. The knowledge-driven models are more interpretable compared to data-driven ones, as they are obtained by theoretically-grounded patterns. Meanwhile, data-driven parameters can extract salient patterns effectively with sufficient data. In some cases, the combination of knowledge- and data-driven trajectories improves the estimation of trait anxiety. These trajectory parameters appear to be reliable predictors of PSA in both real-life and VR settings. In addition, they can be used to predict trait anxiety in real-life scenarios, provided that the model is trained with sufficient data from the VR setting. In this way, the proposed trajectory models learned in VR can mitigate issues related to scarcity and cost when obtaining real-life data from public speaking. In comparison with prior work, Feng *et al.* [12] reported $\rho = 0.30$ in estimating PRPSA using unsupervised transfer learning, while Nirjhar *et al.* [19] obtained $\rho = 0.33$ using the knowledge-based trajectory parameters only. In this paper, we achieved $\rho = 0.39$ and $\rho = 0.34$ (Table 1b) with data-driven and combined models, respectively, suggesting substantial improvement.

The proposed trajectory models can help us better understand causes of PSA during the public speaking encounter and potentially contribute to the prediction of PSA patterns, which can be beneficial to the design of personalized intervention modules [14]. Such estimation of trait anxiety is the first step of designing personalized intervention. For our future work, we plan to conduct third-party temporal annotations of PSA, which will guide the momentary feedback to alleviate PSA. Despite the promising results, our work depicts various limitations. As participants completed 10 public speaking sessions within a short span of time, this might cause habituation to the stimuli. Furthermore, a more sophisticated time-based approach can help design a richer multimodal parameter set. Finally, this work does not include visual information (e.g. facial expression) or third-party annotation, which can potentially enhance the system performance.

7 CONCLUSION

We studied knowledge- and data-driven trajectories of PSA using data from both real-life and VR settings. Parameters obtained from the models exhibit significant association with trait anxiety scores. In addition, models trained on VR settings can reliably predict the trait anxiety in real-life settings. As part of our future work, we will estimate moment-to-moment variation of PSA. We also plan to use more sophisticated data-driven techniques (e.g. transfer learning, few-shot learning) to predict real-life PSA. Finally, we plan to design a real-time PSA detection system, which can facilitate personalized public speaking interventions.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support of the Engineering Information Foundation (EiF #18.02) and the National Science Foundation (#1956021) for this work.

REFERENCES

- [1] Mike Allen, John E Hunter, and William A Donohue. 1989. Meta-analysis of self-report data on the effectiveness of public speaking anxiety treatment techniques. *Communication Education* 38, 1 (1989), 54–76.
- [2] Page L Anderson, Matthew Price, Shannan M Edwards, Mayowa A Obasaju, Stefan K Schmertz, Elana Zimand, and Martha R Calamaras. 2013. Virtual reality exposure therapy for social anxiety disorder: A randomized controlled trial. *Journal of consulting and clinical psychology* 81, 5 (2013), 751.
- [3] Juan Pablo Arias, Carlos Busso, and Nestor Becerra Yoma. 2013. Energy and F0 contour modeling with functional data analysis for emotional speech detection.. In *Interspeech*. 2871–2875.
- [4] Ligia Batrinca, Giota Stratou, Ari Shapiro, Louis-Philippe Morency, and Stefan Scherer. 2013. Cicero-towards a multimodal virtual audience platform for public speaking training. In *International workshop on intelligent virtual agents*. Springer, 116–128.
- [5] Michael J Beatty. 1988. Situational and predispositional correlates of public speaking anxiety. *Communication education* 37, 1 (1988), 28–39.
- [6] Graham D Bodie. 2010. A racing heart, rattling knees, and ruminative thoughts: Defining, explaining, and treating public speaking anxiety. *Communication education* 59, 1 (2010), 70–105.
- [7] Steven Booth-Butterfield and Malloy Gould. 1986. The communication anxiety inventory: Validation of state-and context-communication apprehension. *Communication Quarterly* 34, 2 (1986), 194–205.
- [8] Lei Chen, Gary Feng, Jillian Joe, Chee Wee Leong, Christopher Kitchen, and Chong Min Lee. 2014. Towards automated assessment of public speaking skills using multimodal cues. In *Proceedings of the 16th International Conference on Multimodal Interaction*. 200–203.
- [9] Mathieu Chollet, Torsten Wörtwein, Louis-Philippe Morency, Ari Shapiro, and Stefan Scherer. 2015. 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, 1143–1154.
- [10] Empatica E4. 2021. <https://www.empatica.com/>.
- [11] Florian Eyben, Felix Weninger, Florian Gross, and Björn Schuller. 2013. Recent developments in opensmile, the munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 835–838.
- [12] Kexin Feng, Megha Yadav, Md Nazmus Sakib, Amir Behzadan, and Theodora Chaspari. 2019. Estimating Public Speaking Anxiety from Speech Signals Using Unsupervised Transfer Learning. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 1–5.
- [13] National Collaborating Centre for Mental Health et al. 2013. *Social anxiety disorder: the NICE guideline on recognition, assessment and treatment*. Royal College of Psychiatrists.
- [14] Everlyne Kimani, Timothy Bickmore, Ha Trinh, and Paola Pedrelli. 2019. You'll be Great: Virtual Agent-based Cognitive Restructuring to Reduce Public Speaking Anxiety. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 641–647.
- [15] Gaang Lee, Byungjoo Choi, Changbum Ryan Ahn, and SangHyun Lee. 2020. Wearable Biosensor and Hotspot Analysis-Based Framework to Detect Stress Hotspots for Advancing Elderly's Mobility. *Journal of Management in Engineering* 36, 3 (2020), 04020010.
- [16] Xi Li, Jidong Tao, Michael T Johnson, Joseph Soltis, Anne Savage, Kirsten M Leong, and John D Newman. 2007. Stress and emotion classification using jitter and shimmer features. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, Vol. 4. IEEE, IV–1081.
- [17] James C McCroskey. 1970. Measures of communication-bound anxiety. (1970).
- [18] Angeliki Metallinou, Ruth B Grossman, and Shrikanth Narayanan. 2013. Quantifying atypicality in affective facial expressions of children with autism spectrum disorders. In *2013 IEEE international conference on multimedia and expo (ICME)*. IEEE, 1–6.
- [19] Ehsanul Haque Nirjhar, Amir Behzadan, and Theodora Chaspari. 2020. Exploring Bio-Behavioral Signal Trajectories of State Anxiety During Public Speaking. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1294–1298.
- [20] Virtual Orator. 2021. <https://virtualorator.com/>.
- [21] PACE. 2021. <http://www.stat.ucdavis.edu/PACE/>.
- [22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [23] David P Pertaub, Mel Slater, and Chris Barker. 2001. An experiment on fear of public speaking in virtual reality. *Studies in health technology and informatics* (2001), 372–378.
- [24] Oculus Rift. 2021. <https://www.oculus.com/>.
- [25] Jan Schneider, Dirk Börner, Peter Van Rosmalen, and Marcus Specht. 2015. Presentation trainer, your public speaking multimodal coach. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. acm, 539–546.
- [26] M Iftekhar Tanveer, Samiha Samrose, Raiyan Abdul Baten, and M Ehsan Hoque. 2018. Awe the audience: How the narrative trajectories affect audience perception in public speaking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [27] Azadeh Tavoli, Mahdiyeh Melyani, Maryam Bakhtiari, Gholam Hossein Ghaedi, and Ali Montazeri. 2009. The Brief Fear of Negative Evaluation Scale (BFNE): translation and validation study of the Iranian version. *BMC psychiatry* 9, 1 (2009), 42.
- [28] Helene S Wallach, Marilyn P Safir, and Margalit Bar-Zvi. 2009. Virtual reality cognitive behavior therapy for public speaking anxiety: a randomized clinical trial. *Behavior modification* 33, 3 (2009), 314–338.
- [29] Megha Yadav, Md Nazmus Sakib, Kexin Feng, Theodora Chaspari, and Amir Behzadan. 2019. Virtual reality interfaces and population-specific models to mitigate public speaking anxiety. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–7.
- [30] Megha Yadav, Md Nazmus Sakib, Ehsanul Haque Nirjhar, Kexin Feng, Amir Behzadan, and Theodora Chaspari. 2020. Exploring individual differences of public speaking anxiety in real-life and virtual presentations. *IEEE Transactions on Affective Computing* (2020), 1–1. <https://doi.org/10.1109/TAFFC.2020.3048299>
- [31] Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. 2005. Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* 100, 470 (2005), 577–590.