

# Using Physiological Cues to Determine Levels of Anxiety Experienced among Deaf and Hard of Hearing English Language Learners

Heera Lee  
University of Maryland Baltimore  
County  
Maryland, USA  
heera1@umbc.edu

Varun Mandalapu  
University of Maryland Baltimore  
County  
Maryland, USA  
varunm1@umbc.edu

Jiaqi Gong  
University of Alabama  
Alabama, USA  
jiaqi.gong@ua.edu

Andrea Kleinsmith  
University of Maryland Baltimore  
County  
Maryland, USA  
andreak@umbc.edu

Ravi Kuber  
University of Maryland Baltimore  
County  
Maryland, USA  
rkuber@umbc.edu

## ABSTRACT

Deaf and hard of hearing English language learners encounter a range of challenges when learning spoken/written English, many of which are not faced by their hearing counterparts. In this paper, we examine the feasibility of utilizing physiological data, including arousal and eye gaze behaviors, as a method of identifying instances of anxiety and frustration experienced when delivering presentations. Initial findings demonstrate the potential of using this approach, which in turn could aid English language instructors who could either provide emotional support or personalized instructions to assist deaf and hard of hearing English language learners in the classroom.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

## KEYWORDS

Physiological arousal; Electrodermal activity; English language learners; Deaf and hard of hearing; Anxiety

## ACM Reference Format:

Heera Lee, Varun Mandalapu, Jiaqi Gong, Andrea Kleinsmith, and Ravi Kuber. 2020. Using Physiological Cues to Determine Levels of Anxiety Experienced among Deaf and Hard of Hearing English Language Learners. In *Companion Publication of the 2020 International Conference on Multimodal Interaction (ICMI '20 Companion)*, October 25–29, 2020, Virtual event, Netherlands. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3395035.3425259>

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](mailto:permissions@acm.org).

*ICMI '20 Companion, October 25–29, 2020, Virtual event, Netherlands*

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-8002-7/20/10...\$15.00

<https://doi.org/10.1145/3395035.3425259>

## 1 INTRODUCTION

Deaf and hard of hearing students (D/HH) who come to the US with the aim of strengthening their command of the English language, often find that they need to also learn American Sign Language (ASL) to communicate with instructors and peers through interpreters in their classes. Signs and grammar may differ considerably from the sign language used in their country of origin [4]. Challenges may be faced when attempting to learn both ASL and English at the same time [18]. This extra burden may in turn arouse negative emotions including frustration, anxiety, fear, or stress. However, educators and ASL interpreters, who spend more time with international D/HH students during the semester, may not always notice these issues, as these may be concealed to avoid drawing undue attention or embarrassment.

Researchers have highlighted the benefits of using non-invasive technologies to help identify instances of negative emotional state [15, 19, 27]. Providing greater awareness of students experiencing higher levels of anxiety or frustration can offer potential to English language instructors to provide improved emotional and instructional support to them within a classroom setting [20]. Similarly, this awareness can also prove beneficial for ASL interpreters who may want to offer support to D/HH students within the classroom.

As the first step in this research project, an exploratory study was undertaken to examine the feasibility of using biosensor-based feedback to determine the emotional states of D/HH English language learners (ELLs) when delivering presentations as part of an English language class. In addition to detecting levels of physiological arousal, we integrated other modal data including eye contact with ASL interpreters and the hearing audience, as well as eye gaze (capturing glances to presentation slides and cue cards). These behaviors are typically associated with feelings of anxiety [11]. The main research question addressed in this study is:

- Can biosensor-based feedback and other cues be used to determine at what points do D/HH ELLs experience anxiety when delivering a presentation in front of a hearing audience?

The primary contribution of the work is the development of an innovative approach using physiological arousal to explore the

emotional state of D/HH ELLs. This is one of the first studies examining the feasibility of biosensor-based feedback to determine levels of anxiety faced by individuals with disabilities when delivering presentations within a classroom setting. The approach can be used by English language instructors and ASL interpreters to better understand the emotional states of D/HH ELLs, with a view to providing emotional support or structuring their teaching to address the levels of anxiety/frustration experienced.

## 2 RELATED WORK

As sensing technology has advanced, researchers have attempted to develop interventions to provide feedback on presentation performance (e.g., [9, 10, 12–14, 17, 28, 30, 31, 34]). Examples of data captured include audio data (speech rate, pause) or behaviors (eye contact, body posture, head position).

To help presenters alleviate their anxiety when conducting public speaking activities, Trinh et al.[32] and Kimani et al.[23] proposed an intervention where a virtual coach was used for purposes of rehearsing or when presenting a talk. Takac et al. [29] proposed exposure to a diverse virtual audience when training for public speaking activities. The study identified that exposure to the virtual audience was effective in eliciting public speaking anxiety, as well as habituation through a decrease in response to a stimulus after repeated practice of delivering the presentation. Trinh et al. [33] introduced a system called DynamicDuo, which aimed at improving public speaking anxiety and speaking confidence for non-native English speakers by co-presenting with a virtual presenter in their system. It helps to reduce both the heavy cognitive load associated with presenting in a target language and feelings of stagefright.

As many of the studies cited above focused on presenting within a virtual environment, ecological validity is limited. To measure the anxiety induced through a real world audience, the study described in this paper was conducted in a physical classroom setting where presentation topics were selected by the instructor, and presenters spent time preparing and delivering a presentation as a part of the course.

## 3 METHOD

In our work, we aim to determine the experiences encountered when both presenting independently and co-presenting with other students, and identify instances of anxiety and frustration. Co-presenting is reflective of the type of tasks conducted by language learners within a classroom setting.

### 3.1 Participants

Two participants (1 male (19), 1 female (22), referred to as D11 and D12) were recruited from an English Language Institute in the US where they were enrolled in an English course. They are siblings who hail from Saudi Arabia. Both identified as deaf or hard of hearing from birth. The participants stated that their primary methods of communication were through use of Arabic Sign Language and written Arabic. D11 described experiencing moderate to severe hearing loss (HL) (65dB HL), while D12 highlighted a greater level of loss (85dB HL). Utilizing cochlear implants led to D11 experiencing milder loss (40dB HL), with D12 experiencing moderate levels of loss (41dB HL). Both participants stated that in class, ASL was most

commonly used to communicate with peers and the instructors. However, D11 was able to lip-read spoken Arabic, which could be helpful for communicating with Arabic-speaking students.

There were ten hearing students within the class, alongside D11 and D12. Students were required to deliver a set of presentations in front of the class over the course of the semester. Two ASL interpreters (native English speakers) were employed to translate content from the class, and facilitate communication between D11 and D12 and their peers/instructors. The study protocol was reviewed and approved by the university's Institutional Review Board.

### 3.2 Procedure

E4 wristbands from Empatica Inc. were used to record physiological data when delivering presentations. While the E4 wristband records multiple physiological signals (e.g., blood volume pulse, skin temperature and heart rate), electrodermal activity (EDA) was selected for analysis as the arousal found in the EDA is indicative of anxiety [6]. Participants were observed conducting four classroom-based presentations, while wearing a biosensor wristband to capture EDA data. The presentations were individually delivered or presented with a hearing partner. The first presentation was used to acclimate students to the experimental protocol and wearing the wristband. Analysis was conducted on data collected from the 2nd (individual) and the 3rd (paired) presentations.

After all four presentations were delivered, participants completed a questionnaire called Foreign Language Classroom Anxiety Scale (FLCAS)[22], which has 34-items with a 5-point Likert scale (1=strongly disagree, 5=strongly agree). The items are considered to measure anxiety that arises from an environment where a foreign language is used. Higher scores on these questions correspond to higher levels of anxiety experienced. In addition to the self-assessment questionnaire, a set of semi-structured interviews were conducted after each presentation was delivered. Interviews lasted 15-30 minutes. All interviews were audio recorded and two ASL interpreters took turns to interpret the interview responses or questions into spoken English or ASL.

## 4 ANALYSIS OF DATA AND FINDINGS

### 4.1 Interviews

Participants were asked about their experiences presenting, and to identify anxiety triggers. D11 stated that his anxiety was in part attributed to the personality and teaching style of the instructor, who was thought to be strict. D11 highlighted that his level of nervousness increased when the instructor took notes during his presentation, displayed cue cards indicating time remaining to complete the presentation, and when he could see visible negative facial expressions on the instructor's face (e.g. frowning of brows).

While D11 was aware of his emotional state when delivering presentations in class, D12 stated that she had no time to think about what she was feeling when presenting, as she was busy signing in an unfamiliar language (ASL) while glancing at presentation slides written in English. She shared a negative experience relating to a presentation which she had delivered when in Saudi Arabia. She highlighted that many people were watching her with a "judgmental" look and laughing, which instilled a fear inside her when presenting.

Tonic EDA										
ID	Mean	STD	Peaks	Peaks Count	Min	Min Index	Max	Max Index	Slope	AUC
D11_indiv	-0.18	0.15	0.23	1	-0.26	26	0.23	586	-0.000098	-366.94
D12_indiv	-0.50	0.06	-0.41	1	-0.52	1	0.00	1312	0.000007	-652.66
D11_pair	-0.33	0.05	-0.18	3	-0.35	8	-0.07	2968	0.000011	-991.81
D12_pair	-0.09	0.05	-0.04	4	-0.12	1182	0.10	3068	-0.000023	-262.76

Phasic EDA										
ID	Mean	STD	Peaks	Peaks Count	Min	Min Index	Max	Max Index	Slope	AUC
D11_indiv	0.74	0.28	0.87	104	0.26	2008	1.26	1313	-0.000047	1481.28
D12_indiv	0.90	0.17	1.03	86	0.53	30	1.36	1213	0.000127	1175.06
D11_pair	0.68	0.20	0.75	214	0.35	611	1.15	2885	0.000005	2042.60
D12_pair	0.30	0.09	0.33	263	0.13	2629	0.71	2856	0.000050	914.33

**Table 1: 10 features of tonic and phasic EDA arousal (e.g. mean, standard deviation, minimum and maximum values in a component, locations of minimum and maximum values, mean peak amplitudes, number of peaks, slope, and area under the curve) of D/HH ELLs while delivering an individual and paired presentation in class.**

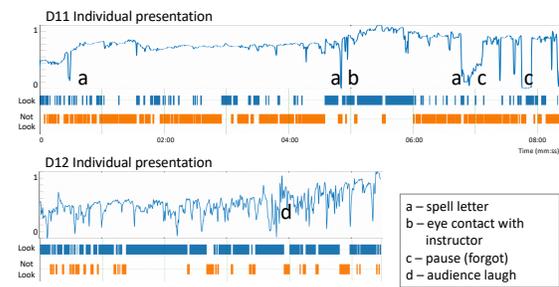
## 4.2 Self-report questionnaire

According to scores from the FLCAS questionnaire [22], D11 and D12 scored 115 and 97 respectively. This placed D11 in the moderate anxiety category (108-144) and placed D12 in the little or no anxiety category (e.g. less than 108) (whole class including hearing ELLs  $n=12$ , mean:  $103.25 \pm 15.33$ ). This total score was analyzed in more detail based upon the items relevant to speech anxiety, fear of failing the class, comfort, and negative attitudes [2]. The most highly anxiety provoking factors identified by D11 related to forgetting presentation content. D11 attempted to memorize content prior to delivering it. D12 was most anxious when comparing herself to other students. Both highlighted that the thought of failing the class was worrisome.

## 4.3 EDA features extraction

Features were extracted from the EDA data collected from the participants. As wearable sensors are prone to various motion artifacts that can adversely impact the EDA data, the data was smoothed using a 1-second Hanning window [1]. We then min-max normalized the data, which mitigated the individual EDA signal differences between subjects [7].

In this study, we extracted two significant components of one-dimensional EDA data that are classified as phasic and tonic components. A tonic component in the signal demonstrates slow changes in the absence of an external stimulus, whereas a phasic component in the signal shows the variations that change quickly over time. We extracted the tonic and phasic components using the non-negative deconvolution algorithm [5]. Ten features of both tonic and the phasic component of EDA signals were extracted and the results of feature extraction for each participant are shown in Table 1. In general, higher mean, peaks (amplitude), and frequent number of peaks indicate anxious states of an individual [6, 24–26].



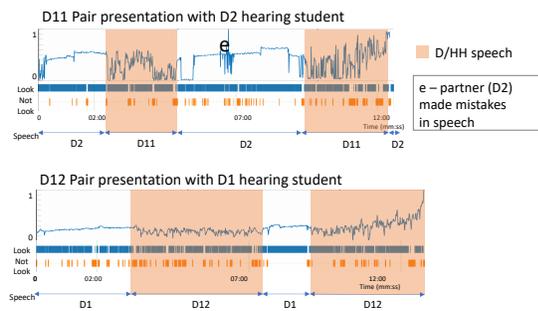
**Figure 1: Normalized EDA signal data and annotation of eye gaze among D/HH students in an individual presentation.**

## 4.4 Eye gaze

We used an annotation tool (BORIS [16]) to manually annotate the audio-video recorded presentations based upon the eye gaze behavior of the participants [11]. As Figure 1 and 2 show, the label ‘Look’ indicates that the participants made eye contact with the audience including classmates, ASL interpreters, and an instructor. On the other hand, the label ‘Not Look’ was used when the participants were observed looking down their notes, or looking either at slides or their partner who they were co-presenting with. Looking at the hearing partner was not counted as looking at the audience.

## 4.5 Corpus Use Cases

**4.5.1 Individual Presentation.** As Figure 1 illustrates, the normalized EDA signal data was synchronized with the eye gaze annotation to explore at what stages during the task that D/HH ELLs experience anxiety. The EDA signal resembles a V shape when D11 was spelling letters of English pronouns, such as name of person or place as finger spelling in ASL (a), made eye contact with an instructor (b), and paused forgetting what to sign next (c). The higher



**Figure 2: Normalized EDA signal data and annotation of eye gaze among D/HH students in a paired presentation.**

frequency in EDA signal of D11 corresponds to the eye contact annotation. On the other hand, D12’s data shows fluctuations in EDA throughout the presentation. The point where the frequency of the EDA signal becomes frequent (d) is matched with the time when the audience including the instructors and classmates were laughing after listening to her speaking about a vicious criminal in her presentation relating to a news report.

**4.5.2 Paired Presentation.** As shown in Figure 2, higher frequencies in EDA signal data were observed among both D/HH students while they were signing ASL (shaded in square). However, the frequency was high for D11 (e) while his partner was presenting, and he was looking at an ASL interpreter translating his partner’s talk into sign language. At that moment, his partner was stuttering because she could not pronounce a certain pronoun in English and then forgot what to say next. She paused for several seconds. In the meantime, the ASL interpreter also paused to interpret for D11, and everyone was looking at the hearing student (D2) for a while. When getting towards the end of the presentation (closer to the Q&A session for the audience), the level of the EDA signal of both D/HH students increases.

## 5 DISCUSSION

We have employed multiple instruments including a self-report questionnaire, individual interviews, a biosensor to identify physiological cues, and audio/video recordings to investigate anxiety among D/HH ELLs when delivering presentations within classroom settings. Using multiple methods allow us to triangulate data from the D/HH ELLs to explore the anxiety among the D/HH ELLs. The most prominent observation from Figure 1 is that the frequency and higher peaks in EDA arousal are linked to the anxiety triggers described in the individual interviews – eye contact with the instructor (b) and a memory of humiliated presentation experience (d). The annotation of eye gaze with the audience, especially with the instructor in the class is matched with the frequency of the EDA arousal with D11 in Figure 1. This provided an insight that the prominent peaks and frequency from the EDA signal data may help instructors and ASL interpreters be able to perceive where the D/HH ELLs feel frustrated and anxious in learning English and ASL process. However, the EDA arousal in Figure 2 is not matched

with the eye gaze annotation even though the D/HH students made consistent eye contact with the audience. It is because they were mostly looking at the ASL interpreter interpreting when their partner was presenting. This suggests that the eye gaze annotation should be divided more specifically based upon whom the participant is looking at (e.g., ASL interpreter, instructor, co-presenter, and the audience) to examine the correlation between the EDA arousal and eye contact with a certain audience. Although D11 reported a higher level of anxiety (115) compared to D12 (97) in the self-report questionnaire, it was not consistent with the EDA data with 10 features. As Table 1, D12 shows higher anxiety than D11 (e.g. more phasic peaks (D11=214, D12=263) in a paired presentation). This may be because the method of self-report is arguably subjective, and the participants can suffer from information recall bias [21].

For individuals who identify as D/HH, prior technologies and studies have focused on developing interventions that can directly help them. For example, Cavender et al. [8] developed the Class-InFocus system to help D/HH students manage multiple visual sources of information from instructor, interpreter, or captions, personal notes by using notification techniques on their system. Al-Khazraji [3] and Wu et al. [35] investigated ways to provide accurate speed and timing of signs. The researchers also improved signing recognition in an application.

However, research to understand the emotional burden among D/HH ELLs taking an English class with other hearing ELLs has not yet been extensively studied within the fields of scientific computing. In this study, we present the points at which the emotional states of the D/HH ELLs vary and identify what triggers their anxiety when undertaking classroom activities. The findings suggest EDA arousal can be a useful indicator to detect the moment where D/HH ELLs feel anxious and frustrated when presenting. With this awareness, English instructors and ASL interpreters in class can spot difficulties faced by D/HH ELLs and prepare tailored materials or structure their teaching to better support them.

## 6 FUTURE WORK

As the next step in the research, signing gestures and speed and accuracy with which these gestures are made, could be explored further to better understand actions made when anxiety is experienced. Additionally, further study should be conducted examining different types of presentation (e.g. planned, impromptu) to identify points at which anxiety is triggered. To validate the findings of the study, the D/HH ELLs who participated in this study and external experts in the fields of speech pathology, sign language, or Deaf community should be invited to provide their input on the findings.

## 7 CONCLUSION

We examined the feasibility of using biosensor-based feedback and behavioral annotation to help instructors and ASL interpreters better identify the emotional states of the D/HH ELLs during a presentation in class. The approach offers promise enabling English language instructors and ASL interpreters to actively approach and provide personalized support to D/HH ELLs. This in turn, will lead to a supportive learning environment.

## REFERENCES

- [1] Lauren B Adamson, Deborah F Deckner, and Roger Bakeman. 2010. Early interests and joint engagement in typical development, autism, and Down syndrome. *Journal of autism and developmental disorders* 40, 6 (2010), 665–676.
- [2] Yukie Aida. 1994. Examination of Horwitz, Horwitz, and Cope’s construct of foreign language anxiety: The case of students of Japanese. *The modern language journal* 78, 2 (1994), 155–168.
- [3] Sedeeq Al-khazraji. 2018. Using Data-Driven Approach for Modeling Timing Parameters of American Sign Language. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. 497–500.
- [4] D Bedoin. 2011. English teachers of deaf and hard-of-hearing students in French schools: Needs, barriers and strategies. *European Journal of Special Needs Education* 26, 2 (2011), 159–175.
- [5] Mathias Benedek and Christian Kaernbach. 2010. Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology* 47, 4 (2010), 647–658.
- [6] Wolfram Boucsein. 2012. *Electrodermal activity*. Springer Science & Business Media.
- [7] Jason J Braithwaite, Derrick G Watson, Robert Jones, and Mickey Rowe. 2013. A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology* 49, 1 (2013), 1017–1034.
- [8] Anna C Cavender, Jeffrey P Bigham, and Richard E Ladner. 2009. ClassInFocus: enabling improved visual attention strategies for deaf and hard of hearing students. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*. 67–74.
- [9] Lei Chen, Chee Wee Leong, Gary Feng, and Chong Min Lee. 2014. Using multimodal cues to analyze mla’14 oral presentation quality corpus: Presentation delivery and slides quality. In *Proceedings of the 2014 ACM workshop on multimodal learning analytics workshop and grand challenge*. ACM, 45–52.
- [10] Mathieu Chollet and Stefan Scherer. 2017. Assessing public speaking ability from thin slices of behavior. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. IEEE, 310–316.
- [11] Mathieu Chollet, Torsten Wörtwein, Louis-Philippe Morency, and Stefan Scherer. 2016. A multimodal corpus for the assessment of public speaking ability and anxiety. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. 488–495.
- [12] Keith Curtis, Gareth JF Jones, and Nick Campbell. 2015. Effects of good speaking techniques on audience engagement. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. 35–42.
- [13] Ionut Damian, Chiew Seng Sean Tan, Tobias Baur, Johannes Schöning, Kris Luyten, and Elisabeth André. 2015. Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques. In *Proceedings of the 33rd annual ACM conference on Human factors in computing systems*. ACM, 565–574.
- [14] Fiona Dermody and Alistair Sutherland. 2016. Multimodal system for public speaking with real time feedback: a positive computing perspective. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. 408–409.
- [15] Elena Di Lascio, Shkurta Gashi, and Silvia Santini. 2018. Unobtrusive assessment of students’ emotional engagement during lectures using electrodermal activity sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–21.
- [16] Olivier Friard and Marco Gamba. 2016. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution* 7, 11 (2016), 1325–1330.
- [17] Michelle Fung, Yina Jin, RuJie Zhao, and Mohammed Hoque. 2015. ROC speak: semi-automated personalized feedback on nonverbal behavior from recorded videos. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 1167–1178.
- [18] CL Garberoglio, JL Palmer, and S Cawthon. 2019. Undergraduate enrollment of deaf students in the United States.
- [19] Shkurta Gashi, Elena Di Lascio, and Silvia Santini. 2019. Using unobtrusive wearable sensors to measure the physiological synchrony between presenters and audience members. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–19.
- [20] Tammy Gregersen et al. 2020. Dynamic properties of language anxiety. *Studies in Second Language Learning and Teaching* 10, 1 (2020), 67–87.
- [21] Javier Hernandez, Ivan Riobo, Agata Rozga, Gregory D Abowd, and Rosalind W Picard. 2014. Using electrodermal activity to recognize ease of engagement in children during social interactions. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 307–317.
- [22] Elaine K Horwitz, Michael B Horwitz, and Joann Cope. 1986. Foreign language classroom anxiety. *The Modern language journal* 70, 2 (1986), 125–132.
- [23] Everlyne Kimani and Timothy Bickmore. 2019. Addressing Public Speaking Anxiety in Real-time Using a Virtual Public Speaking Coach and Physiological Sensors. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. ACM, 260–263.
- [24] Heera Lee and Andrea Kleinsmith. 2019. Public Speaking Anxiety in a Real Classroom: Towards Developing a Reflection System. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, LBW2417.
- [25] Heera Lee, Varun Mandalapu, Andrea Kleinsmith, and Jiaqi Gong. 2020. Distinguishing Anxiety Subtypes of English Language Learners Towards Augmented Emotional Clarity. In *International Conference on Artificial Intelligence in Education*. Springer, 157–161.
- [26] Janick Naveteur and Esteve Freixa I Baque. 1987. Individual differences in electrodermal activity as a function of subjects’ anxiety. *Personality and individual differences* 8, 5 (1987), 615–626.
- [27] Ming-Zher Poh, Nicholas C Swenson, and Rosalind W Picard. 2010. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE transactions on Biomedical engineering* 57, 5 (2010), 1243–1252.
- [28] 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.
- [29] Marcel Takac, James Collett, Kristopher J Blom, Russell Conduit, Imogen Rehm, and Alexander De Foe. 2019. Public speaking anxiety decreases within repeated virtual reality training sessions. *PLoS one* 14, 5 (2019), e0216288.
- [30] M Iftexhar Tanveer, Emy Lin, and Mohammed Ehsan Hoque. 2015. Rhema: A real-time in-situ intelligent interface to help people with public speaking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, 286–295.
- [31] M Iftexhar Tanveer, Ru Zhao, Kezhen Chen, Zoe Tiet, and Mohammed Ehsan Hoque. 2016. Automanner: An automated interface for making public speakers aware of their mannerisms. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 385–396.
- [32] Ha Trinh, Reza Asadi, Darren Edge, and T Bickmore. 2017. RoboCOP: A Robotic Coach for Oral Presentations. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 27.
- [33] Ha Trinh, Lazlo Ring, and Timothy Bickmore. 2015. Dynamicduo: co-presenting with virtual agents. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1739–1748.
- [34] Torsten Wörtwein, Mathieu Chollet, Boris Schauer, Louis-Philippe Morency, Rainer Stiefel, and Stefan Scherer. 2015. Multimodal public speaking performance assessment. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. 43–50.
- [35] Jiaxiang Wu, Jian Cheng, Chaoyang Zhao, and Hanqing Lu. 2013. Fusing multimodal features for gesture recognition. In *Proceedings of the 15th ACM on International conference on multimodal interaction*. 453–460.