SmartAudience: A mobile APP designed for presentation training

Zhihe Zhao
Duke University
zz208@duke.edu

Wanli Wang
Duke University
ww165@duke.edu

ABSTRACT

Oral presentation skills are central to physician-physician communication;however, little is known about how these skills are learned. These skills are required by people whether they are expected to give presentations at conferences, symposia or other meetings. Because of the necessity for theses kills, we need instructions and guidance in preparing, organizing and delivering oral presentations. Besides the essence of communicating in professional life, people regularly underestimate the amount of time professionals spent on presentation and other forms of communication. Therefore, specific attention is required as to the design of educational products to develop oral presentation performance. We design an APP named SmartAudience which is aimed to help users people improve their presentation skills from all aspects.

CCS CONCEPTS

• **Human-centered computing** → **Graphical user interfaces**.

KEYWORDS

Mobile computing, HCI, deep learning

INTRODUCTION

Oral presentations are becoming an important part, especially in the university environment [16]. For example, the importance of teaching students presentations for language learning, the ability to communicate is the most important goal that communicative language teaching aims to reach. It is to be able to operate effectively in the real world. Students need a lot of opportunity to practice language in situations which encourage them to communicate their needs, ideas and opinions. For successful oral communication students need a thorough in-depth instruction and practice. The best practice is to give oral presentations. Oral presentation skills are essential for employability and true

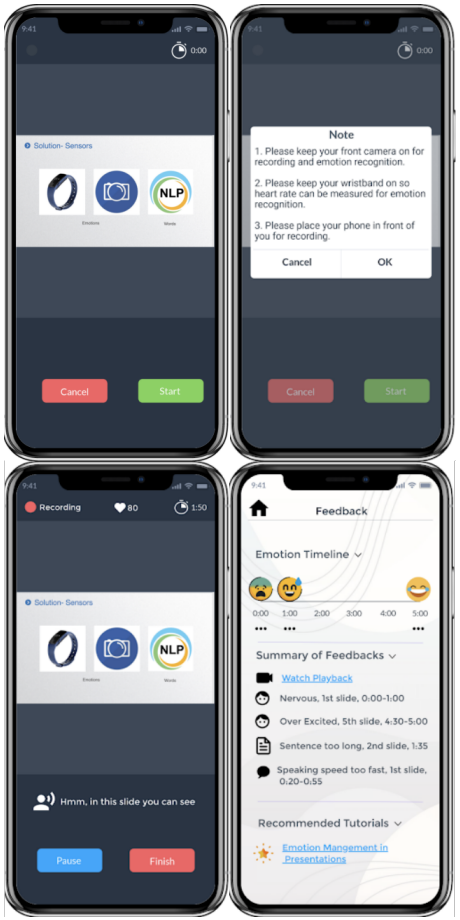

Figure 1: The interface of SmartAudience.

academic study as they lead students to enter into debate and sustained reasoning [16][13]. However, professionals from several domains emphasize that graduates often lack the competence to speak in public or give a presentation.

Oral presentation skills are central to physician-physician communication; however, little is known about how these skills are learned [4]. These skills are required by people whether they are expected to give presentations at conferences, symposia or other meetings. Because of the necessity for these skills, we need instructions and guidance in preparing, organizing and delivering oral presentations[1]. Our product SmartAudience as shown in Fig. 1 is aimed to help users (especially students) improve their presentation skills from all aspects. We embed deep learning algorithms in SmartAudience in order to effectively recognize presenters' basic performance such as speaking speed and advanced features of performance such as emotional expressions variations in real time. These recognition results would be useful for further analysis on the general performance of the presenters.

RELATED WORK

Emotions in presentations. According to Woodrow's work [1], anxiety reactions can be categorized as reflecting worry or emotionality. During the presentation, emotionality refers to physiological reactions, such as blushing or racing heart, and behavioural reactions, such as stammering and fidgeting. Thus, one of the functionality of our product is to accurately recognize the users' emotions and record them in a timeline for reference. Based on this, we are able to provide further tips or advice on corresponding emotion problems of the presenter.

Tools & Metrics in presentations. In order to improve users' presentations, prior work has investigated how to recognize emotions during people's presentations, peer assessment of presentations and metrics on presentation performance. Regarding emotion recognition during presentations, [7][15][14][5] review and design emotion recognition based on video streams and speech. Work in [5] reviews emotion recognition from speech signals, which is an important but challenging component of Human-Computer Interaction (HCI). About the presentation assessment, Grez's work [2] gives a detailed analysis. The study focuses on the agreement between professional assessment and self- and peer assessment of oral presentation skills.

Similar presentation trainer products. Here we also present similar products, which can be divided into two categories: multimodal presentation trainer [3][11][10] and virtual-reality based presentation trainer[12][6][9]. Work in [11] mainly uses kinect features and speech features to generate a general performance score. However, works like this require a huge cost on the devices, which may become a barrier to users. Regarding virtual-reality based presentation trainers, a representative work is [12]. In this work, the authors added two new features to the PT: a VR real-time feedback module and a post-practice histogram and video analysis. However, virtual reality based products can hardly support users to practice for a long time because of the drawback of VR itself. Users may feel dizzy if

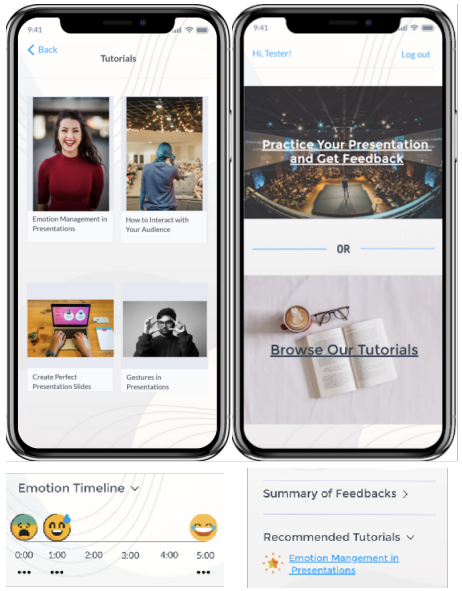

Figure 2: The tutorial page, main page as well as emotion timeline analysis interfaces of SmartAudience.

wearing the equipment for more than half an hour. In SmartAudience, we fuse multiple sensors (wrist bands, camera speech sensor) together to collect users' presentation related data, while developing the APP which is portable and easy for users to get feedback. We also provide recommendations and advice after analyzing users' presentation performance.

DESIGN AND IMPLEMENTATION

Here we list all the features designed in SmartAudience and some preliminary implementations on this product.

Login, sign up. This feature is a commonly seen feature of many apps.

Browse Tutorials. Users can choose to browse all tutorials provided by the app. The major way of providing tutorials to uses in our app is to recommend tutorials to them based on the feedback. However, the users may want to check more tutorials apart from recommended tutorials. There is also a concern of respecting the user's autonomy and avoiding deception. Since the user may have their own needs for tutorials and may not be satisfied with the recommended tutorials. Therefore, giving them full access to tutorials is necessary.

Recording. The app will record the user's video(facial expression), the user's screen(with slides), the user's words(recognized by NLP) and the user's heart rate(wristband). The recordings will later be used for playback, the user can use the playback to check their performance and locate their problems.

Emotion and Word Recognition. Based on the video of the user and the heart rate of the user, the app will detect the user's emotions. NLP technology will be used to recognize the user's words.

Smart Advisor. The app will also act as a smart advisor and provide feedback on the presentation and recommendations on tutorials.

Emotion Timeline. A timeline that marks the time of major fluctuations in emotions and extreme emotions will be provided as part of the feedback. One of the problems of having an actual audience to give feedback or watching recorded presentations is that when the presentation gets long, it is very likely that the audience and the presenter would forget how some parts of the presentation goes. To avoid this problem, the designed emotion timeline will give the user both a general view of the presentation and a detailed tracking of emotion changes. The interface of this as well as the tutorial page are shown in Fig. 2.

Feedback. Feedback on both the user's emotions and the user's words will be provided. The app will provide when and where problems happened as well.

Recommendation of Tutorials. Tutorials will be recommended to users based on the feedback. This would make studying from tutorials most efficient. Since the problem of hard to find the tutorial to check will be solved, the user can use the tutorial to learn about the exact part that they need to improve on.

Regarding the implementation of facial emotion recognition, we used Pytorch[8] to implement and

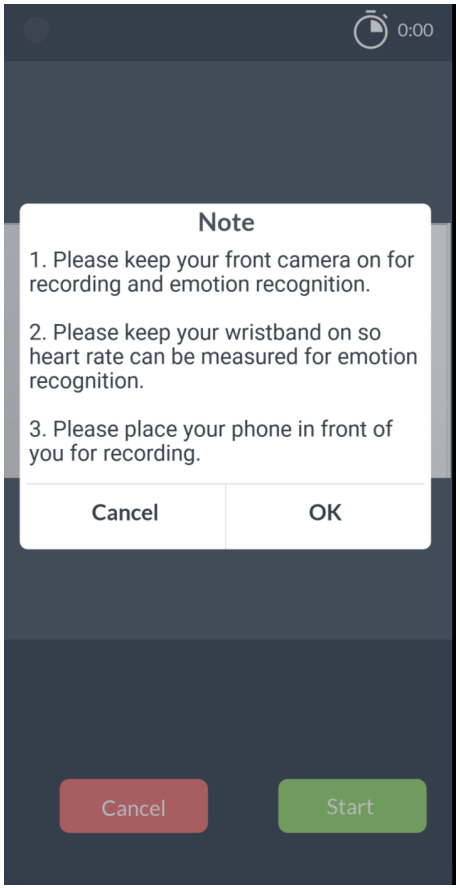

Figure 3: Ethical consideration of SmartAudience.

train all the models using stochastic gradient descent with momentum. The dataset we used is FER2013 which has seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The training set consists of 28,709 examples and the public test set consists of 3,589 example. We used ResNet20 as the baseline. The training was executed on a Linux platform integrated with a 11 GB-memory GPU (Nvidia GeForce RTX 2080 Ti). The training was performed for 20 epochs with mini-batches of size 14. In each batch, the numbers of samples for each class were restricted to be the same. The learning rate was set to 0.01.

ETHICAL CONCERN

Privacy. Our app will collect some data from the user to make emotion recognition and presentation feedback. The user's video and voice will be recorded for later playback. Meanwhile, from the user's video, we would detect their facial expressions and use them for emotion recognition. A wristband will be recommended for users to wear during the presentation practice. The user's voice will also be analyzed by NLP algorithms. To protect user's privacy, we have two major ways.

The first way is to get the user's consent. As Fig.3 shows, we would give a notification before each time the user clicks the 'start' button to start a practice. The second way is regarding user data management, we plan to store the user's data locally, especially the video of the user and only store the user's practice history and feedback in the backend. We also plan to give the user a way to request erasing all personal data, possibly by adding an option or a button in the main page.

Deception and Autonomy. In order to avoid deception or violate user's autonomy, we tried to make sure that our users can get enough information to make their own judgement and decision. In playback: users can always watch their recorded presentation again to double check the pointed out problems and decide what to do about them. Timestamps and jump buttons are added to help them locate the problems. In tutorials: besides the recommended tutorials, users can always browse all tutorials and decide what they would like to learn.

CONCLUSION & FUTURE WORK.

We design a smart presentation training app named SmartAudience which acts as a smart audience. The user can practice presentations in the app, which can analyze the user's performance and give feedback.The app can also recommend tutorials to users based on their previous presentation performance from the aspects of emotion management, grammar as well as speech performance. In the future, we consider add more functionalities into SmartAudience, for example, providing augment reality based interacting tools which can show the performance tips more vividly for users. Besides, we plan to further improve the recognition accuracy by integrating more powerful deep neural networks into the APP.

REFERENCES

[1] Page L Anderson, Elana Zimand, Larry F Hodges, and Barbara O Rothbaum. 2005. Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and anxiety* 22, 3 (2005), 156–158.

[2] Luc De Grez, Martin Valcke, and Irene Roozen. 2012. How effective are self-and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learning in Higher Education* 13, 2 (2012), 129–142.

[3] 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.

[4] Richard J Haber and Lorelei A Lingard. 2001. Learning oral presentation skills. *Journal of general internal medicine* 16, 5 (2001), 308–314.

[5] Ruhul Amin Khalil, Edward Jones, Mohammad Inayatullah Babar, Tariqullah Jan, Mohammad Haseeb Zafar, and Thamer Alhussain. 2019. Speech emotion recognition using deep learning techniques: A review. *IEEE Access* 7 (2019), 117327–117345.

[6] Bas Kollöffel and Kirsten Olde Heuvel. 2020. VIRTUAL REALITY TRAINING OF PRESENTATION SKILLS: HOW REAL DOES IT FEEL? A MIXED-METHOD STUDY.. In *48th SEFI Annual Conference on Engineering Education, SEFI 2020*. University of Twente, 252–261.

[7] X Ochoa, F Domínguez, B Guamán, R Maya, G Falcones, and J Castells. 2018. The RAP system: automatic feedback of oral presentation skills using multimodal analysis and low-cost sensors (pp. 360–364).

[8] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703* (2019).

[9] David-Paul Pertaub, Mel Slater, and Chris Barker. 2002. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators & Virtual Environments* 11, 1 (2002), 68–78.

[10] Vikram Ramanarayanan, Chee Wee Leong, Lei Chen, Gary Feng, and David Suendermann-Oeft. 2015. Evaluating speech, face, emotion and body movement time-series features for automated multimodal presentation scoring. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. 23–30.

[11] 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*. 539–546.

[12] Jan Schneider, Gianluca Romano, and Hendrik Drachsler. 2019. Beyond reality—Extending a presentation trainer with an immersive VR module. *Sensors* 19, 16 (2019), 3457.

[13] Victor N Shaw. 2001. Training in presentation skills: an innovative method for college instruction. (2001).

[14] Ashish Tawari and Mohan Manubhai Trivedi. 2010. Speech emotion analysis: Exploring the role of context. *IEEE Transactions on multimedia* 12, 6 (2010), 502–509.

[15] Kulawadee Yamkate and Charatdao Intratat. 2012. Using video recordings to facilitate student development of oral presentation skills. *Language Education in Asia* 3, 2 (2012), 146–158.

[16] Slađana Živković. 2014. The importance of oral presentations for university students. *Mediterranean Journal of Social Sciences* 5, 19 (2014), 468–468.