EcoScript: A Real-Time Presentation Supporting Tool using a Speech Recognition Model

Eunyeoul Lee[1*], Eunseo Yang[1*], Jinyoung Huh[2], Uran Oh[2†]
Artificial Intelligence Convergence[1], *Computer Science and Engineering*[2]
Ewha Womans University, Seoul, Republic of Korea
{e.lee, y.eunseo, hjy0505}@ewhain.net, uran.oh@ewha.ac.kr

***Abstract*—Delivering scripted content while maintaining ideal vocal qualities and engaging with an audience during a presentation is challenging. While recent tools use advanced technologies such as VR and AR to support presentations, these can be cumbersome in real presentation environments. To address these issues, we conducted a preliminary survey with 81 participants to identify the desired features when using PowerPoint Presenter Mode. Based on the survey results, we propose EcoScript, a web-based presentation-supporting tool that focuses on real-time voice analysis and script tracking. A user study with 20 participants showed that EcoScript significantly enhances time management, voice modulation, and overall satisfaction, indicating that it effectively resolves common challenges encountered in the presentation process. These results suggest a direction for the future of presentation technology, promoting the development of more accessible and practical presentation support tools.**

***Index Terms*—Public Speaking, Presentation Tool, Speech Recognition, User Experience Design**

I. Introduction

Effective oral communication skills are essential across personal, academic, and workplace settings [1], [2]. These skills extend beyond merely delivering information; they encompass sharing ideas and persuading others, crucial for influencing stakeholders and advancing organizational goals [3]–[7].

However, the presentation process entails numerous challenges, such as time management, effective content delivery, and interaction with the audience [8], [9]. To overcome these issues, various technological support tools have been researched and developed, including automatic oral presentation feedback systems [10]–[12], and real-time presentation feedback using AR glasses [13], [14].

Yet, those AR and VR-based presentation support tools can burden presenters. The weight and discomfort of these devices may hinder user experience [15], [16] and limit their practical effectiveness in real presentation situations. Additionally, presenters need to focus on their presentation content and audience reactions, and having to manipulate additional technical equipment can instead add to their burden in the presentation situation.

This study, recognizing the challenges in current presentation technologies, proposes a new approach to maximize the presenter's convenience and enhance the quality of presentations. To inform the development of our prototype, we surveyed 81 university students to understand their preparation and behavior patterns during presentations and the inconveniences they encounter. The survey found that over 80% of respondents pre-write and use scripts during presentations. They often utilize features like presenter notes, elapsed time monitoring, and next-slide preview. The main reason for using presenter notes was fear of public speaking and difficulty memorizing long scripts. Additionally, participants showed significant interest in receiving voice feedback during presentations.

Leveraging insights from the formative survey, we focused on developing a web-based prototype that utilizes voice recognition technology to track the presentation script and provide real-time analysis of the presenter's volume, pitch, and speaking rate through graphs. Our system, EcoScript, enables presenters to track the script using only their voice, significantly enhancing convenience during presentation execution.

As for the evaluation, we validated the effectiveness of EcoScript through a user study involving 20 participants, who reported significant improvements in presentation convenience. This success underscores the advantages of employing advanced voice recognition technologies in presentations, which not only enhance the convenience for the presenter but also increase process efficiency. EcoScript intelligently progresses the presentation based on the completion of the script on the current slide or automatically scrolls through text that exceeds the visible window area. Additionally, the real-time voice analysis graph displayed by EcoScript helped participants maintain a steady voice throughout their presentations.

Based on these results, this study has made important contributions to the development of presentation skills and tools by deeply understanding the difficulties faced by presenters during the presentation process and proposing a technological approach to address them. Advancing this research can create an environment where presenters can concentrate better and communicate their messages more effectively.

II. Related Work

A. Automatic Oral Presentation Feedback System

The automatic oral presentation feedback system is a system designed to improve oral presentation skills. It analyzes various modalities of information obtained from the presenter during the presentation and provides feedback [17]–[19]. These

*First co-authors, †Corresponding author

systems have mainly been studied to automatically evaluate the presentation or interview skills of students or interviewees and help them practice presentations. For this purpose, they analyze the user's voice and non-verbal behaviors such as gestures and facial expressions to provide feedback [10], [13], [18]. Since presentations are conducted through language, most of the research on automatic oral presentation feedback systems deals with voice analysis [17]–[20]. Our study also follows the current trend of automatic oral presentation feedback systems, providing real-time graphs of common voice analysis features such as volume, pitch, and speaking rate. In addition, this study has adopted the function from previous studies to highlight the part that the presenter has already read, allowing the presenter to make eye contact with the audience and easily return to the original position in the script after the interaction [21]. This helps maintain the continuity of the presentation and allows the presenter to focus more on their presentation. Furthermore, we propose a UI that, instead of displaying the entire script for each slide, presents the script divided into multiple cue cards based on each subheading within a slide. This not only reduces the presenter's cognitive load [22] but also supports seamless presentation continuation while interacting with the audience.

B. Real-Time Feedback During Presentations

Real-time feedback has been demonstrated to significantly impact the development of presentation skills, supporting presenters in real-time adjustments and improvements [10], [12], [13], [23]–[25]. These studies emphasized the importance of non-verbal signals during presentations and explored methods to improve them through a technological approach. For example, Schneider et al. developed a real-time feedback tool to provide users with feedback on their non-verbal communication and demonstrated that this tool could contribute to the improvement of presentation skills [10]. Similarly, Tanveer et al. utilized Google Glass to provide real-time feedback on the presenter's voice volume and speaking rate, which could increase awareness of non-verbal behaviors during presentations [13]. Inspired by these studies, our system provides real-time voice feedback during presentations, helping presenters monitor and immediately adjust their speech volume, pitch, and rate. This approach not only enhances awareness of non-verbal communication but also contributes to more confident and effective presentation delivery.

III. Formative Study: Online Survey

In this section, we describe the design approach for the AI-based presentation aid tool prototype, along with its user interface and key functionalities.

A. Survey Design

We conducted a 12-day survey using Google Forms with 81 participants recruited from university communities. Of these participants, 43 were between 18 and 24, while the rest were aged between 25 and 34. The survey required participants to have experience with preparing and delivering PowerPoint presentations and included 28 questions, taking approximately 5 minutes to complete. Participants were eligible for coffee vouchers through a random draw.

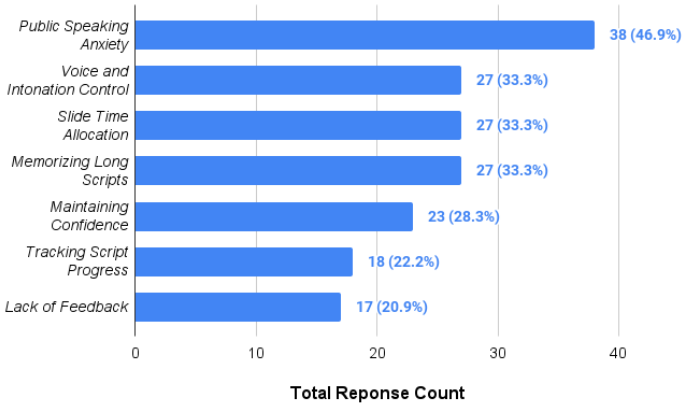

Fig. 1: Challenges in Presentation. Multiple answers were allowed (N = 81).

B. Results

As seen in Figure 1, a survey on the difficulties of presenting found that 46.9% of participants identified fear and anxiety about public speaking as their primary challenge, with a tendency to rely on scripts heavily. Participants, constituting 33.3% reported difficulties in remembering their prepared scripts during delivery. Thirty-three point three percent of participants reported difficulty maintaining appropriate vocal characteristics during presentations. They cited nervousness as hindering their ability to focus on aspects such as speaking pace, tone, and vocal clarity. Furthermore, 22.2% of participants faced difficulty tracking their place in the script after attempting eye contact and gestures to engage the audience, especially with long scripts that required scrolling. Lastly, 20.9% of participants indicated that the lack of instant overall feedback was a challenge. Participants also experienced difficulty in time management.

Hence, presenters' challenges can be condensed into four main categories: anxiety, lack of real-time vocal feedback, content tracking burdens, and time management issues. These findings emphasize the necessity for real-time automated support software to ease presenter challenges.

IV. EcoScript

Based on the identified presenter's needs, we propose the web-based software prototype EcoScript to facilitate time management, real-time voice feedback, and content tracking. The overall user interface is depicted in Figure 2. We utilized React [26] for frontend development and FastAPI [27] for backend server construction.

A. Presentation Slide Control and Presenter Mode Access

The current slide being presented is displayed in the Figure 2-(c) area, allowing presenters to review pre-stored scripts and slides. They can navigate to the next slide using the button in Figure 2-(d). The button in Figure 2-(b) allows the user to start

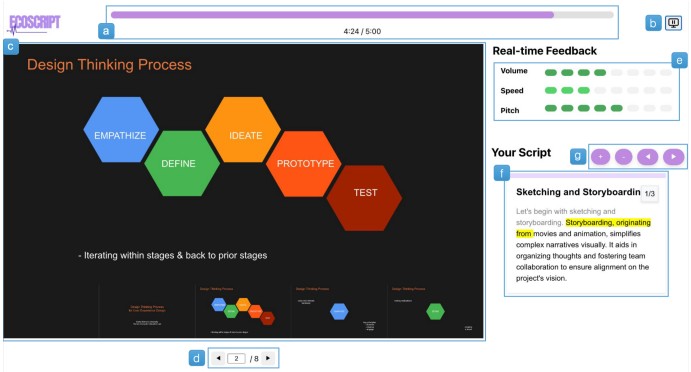

Fig. 2: System Design: (a) Timer progress bar area, (b) Presentation mode start/stop button, (c) Current slide area and preview board, (d) Slide page control buttons, (e) Real-time voice feedback graph area, (f) Cue cards and script tracking area, (g) Cue card font and page control button area.

or stop functions such as the timer, script tracking, and voice analysis and feedback.

B. Countdown-Style Timer

As shown in Figure 2-(a), EcoScript employs a countdown approach rather than the conventional count-up method used in other presentation software. Presenters can intuitively monitor the time elapsed relative to the designated presentation time through a progress bar.

C. Real-Time Voice Feedback

The key elements of vocal feedback we focused on are pitch, volume, and speed. All vocal metrics are updated in real-time on the feedback area as bars. Voice pitch is manually adjusted within the normal range of 50Hz to 250Hz for males and 120Hz to 350Hz for females [28], [29]. To analyze the presenter's voice features, during the presentation, we used a Praat Python Library [30]. This library enables the analysis of the voice volume and pitch of the speaker's voice. Speed is calculated at 0.5 seconds per word, measured by the words per second analyzed through the Google Cloud Speech to Text(STT) API. The area in Figure 2-(e) shows real-time feedback graphs of vocal metrics like volume, speed, and pitch. The bars of the graph are divided into a total of 8 sections to represent the degree of numerical values. If sections 1, 2, 7, and 8 of the graph are colored, they indicate abnormal values and are displayed in red. Sections 3 through 6, when colored, represent optimal values and are shown in green.

D. Script Tracker

The presenter's voice is converted into text via the Google Cloud STT API [31] and matched in real-time with the pre-saved script. As the spoken words align with the script, the corresponding text is highlighted in the 'Your Script' area of Figure 2. Words already read are grayed out, and the next three words are highlighted in yellow. The scroll adjusts automatically to keep the referenced line visible. Once all the text on a cue card is read, it advances to the next cue card automatically, allowing the presenter to focus on the presentation without manual navigation. If the presenter speaks words not in the script (such as answering questions), the system does not automatically track those. Additionally, EcoScript improves script management efficiency by breaking the script into multiple subheadings instead of writing everything on one slide. This enhances the presenter's ability to recall content easily, increasing both the flexibility and focus of the presentation. Presenters can navigate to the next cue cards in case of voice recognition errors. These controls are located in Figure 2-(g).

V. Evaluation

We evaluated Ecoscript's convenience, effectiveness, and user experience compared to Microsoft PowerPoint through a within-subject study with 20 participants (12 female, 8 male). Since most participants in the formative study used MS PowerPoint, it was chosen as the comparison platform.

A. Participants

Participants averaged 23.8 years old, mainly university students, including some working professionals and graduate students. Selection criteria included no visual or auditory impairments, prior experience with PowerPoint's presenter mode, and familiarity with preparing presentation scripts.

B. Experiment Settings

We used Tobii Pro Glasses 3 wearable eye tracker to monitor eye movements, focusing on screen areas and audience eye contact during presentations. Two 5-minute presentations based on Stanford University's 2022 HCI course (CS147) [1] content were prepared, covering "Early Stage Prototyping" and "Design Thinking Process for User Experience Design." Two laptops were used for the presentations: one with EcoScript and the other with Microsoft PowerPoint. To create a realistic presentation environment, the presentations were projected onto a screen, and audience pictures were placed in front.

C. Procedure

Participants received a 5-minute presentation slide and script PDF in advance to familiarize themselves with the content. Upon arrival, they received a brief tutorial on EcoScript and Microsoft PowerPoint presenter mode.

Before starting, participants were instructed to maximize eye contact with the audience, maintain an ideal voice level, and keep within a five-minute limit. Presentation order was counterbalanced. Participants were immediately asked to rate their experience using a 7-Likert scale to gather their opinions on each tool.

VI. Results

As for the analysis, we used the Wilcoxon signed-rank test and gaze tracking logs to evaluate the differences between using EcoScript and Microsoft PowerPoint. The results are shown in Figure 3 and Figure 5.

[1]https://hci.stanford.edu/courses/cs147/2022/au/

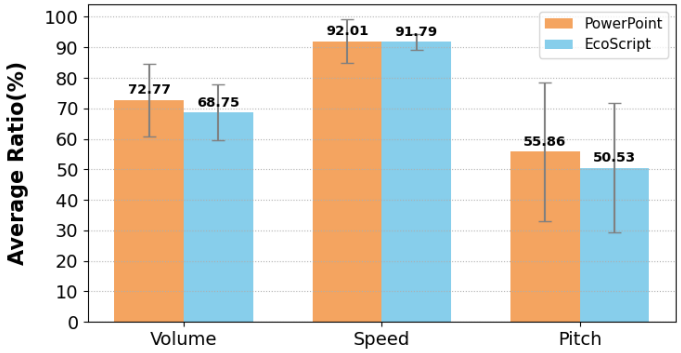

Fig. 3: Average Ratio of Ideal Voice Features (Volume, Speed, and Pitch) with Standard Deviation

A. Ideal Ratio for Volume, Speed and Pitch

Based on the collected logs, we assessed the duration where each of the following metrics was considered ideal (see Section IV-C) over the entire presentation time: volume, speed, and pitch. As a result, we found that the feedback was most helpful in maintaining the ideal speech rate (i.e., speed) followed by volume and pitch as shown in Figure 3.

B. Audience Interaction: Eye Contact

Figure 4 shows the gaze tracking logs as a heatmap during the experiment when presenters looked at the audience's photo. Areas glanced at once are green, while those looked at repeatedly turn red.

Both PowerPoint and EcoScript had similar numbers of eye contacts, but EcoScript maintained these contacts for longer. Presenters were advised to maximize eye contact. Figure 4-(a) shows PowerPoint participants made frequent brief glances away, while EcoScript in Figure 4-(b) allowed longer sustained eye contact. This was because EcoScript enabled quick identification of necessary script segments in real-time, helping presenters focus more on the audience and maintain engagement.

C. Efficiency in Presentation Time Allocation (Time)

Participants rated EcoScript's time management efficiency significantly higher than PowerPoint, with an average rating

(a) PowerPoint Presentation Eye Tracker Heatmap

(b) EcoScript Presentation Eye Tracker Heatmap

Fig. 4: Eye Tracking Analysis Comparison of P8

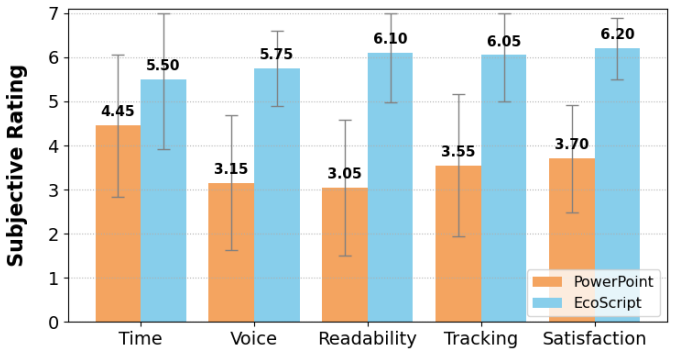

Fig. 5: Average Scores for PowerPoint vs EcoScript (Evaluation metrics: Time, Voice, Readability, Tracking, Satisfaction)

of 5.5 (SD=1.573) for EcoScript and 4.45 (SD=1.605) for PowerPoint ($p < .001$), as shown in the 'Time' bar in Figure 5.

EcoScript received positive feedback for its visual aids. In contrast, PowerPoint participants reported time management difficulties due to needing to check the time, which added mental strain and disrupted their flow. This underscores the importance of a tool's time display features for maintaining focus and effective time management.

D. Stability of the Presenter's Voice (Voice)

EcoScript was rated higher for voice stability, with an average score of 5.75 (SD=0.851) compared to PowerPoint's 3.15 (SD=1.531) ($p < .001$), indicating more consistent voice quality with EcoScript, as shown in the 'Voice' bar in Figure 5.

Participants gave positive feedback on EcoScript, particularly its real-time voice monitoring. They appreciated the visual feedback for reducing nervousness and boosting self-confidence. However, some mentioned challenges with the rapid movement of the feedback bar.

E. Ease of Reading Long Scripts (Readability)

In terms of script readability, EcoScript was preferred for script readability, scoring 6.10 (SD=1.119) compared to PowerPoint's 3.05 (SD=1.538) ($p < .001$), likely due to its cue card format. See 'Readability' bar in Figure 5.

Most Participants appreciated the cue card format and voice-activated screen advancement, which allowed them to focus on the audience without controlling the script manually. This setup facilitated appropriate pauses with automatically advancing small paragraphs.

Moreover, the combination of automatic scrolling and visual highlighting made it easier to manage lengthy scripts, enabling smooth content delivery.

F. Ease of Tracking the Script Position (Tracking)

EcoScript scored higher for ease of tracking script position, with an average of 6.05 (SD=1.050) compared to MS PowerPoint's 3.55 (SD=1.605) ($p < .001$), indicating a more user-friendly interface. See 'Tracking' bar in Figure 5.

With MS PowerPoint, presenters manually checked their script position, often losing their place while making eye contact. In contrast, EcoScript's automatic tracking and highlighting features reduced this stress, aiding smooth presentation flow.

Especially, presenters appreciated quickly locating their position in the script after eye contact. However, challenges included estimating remaining time without knowing the number of paragraphs left and needing adjustments for speech recognition errors.

G. Overall Satisfaction (Satisfaction)

The overall satisfaction with EcoScript was significantly greater, as reflected by an average score of 6.20 (SD=0.696) compared to PowerPoint's 3.70 (SD=1.218) ($p < .001$). The enhancements offered by EcoScript, including real-time feedback and automatic scrolling, contributed to this result, as depicted in the 'Satisfaction' bar in Figure 5.

Additionally, participants expressed a demand for features such as additional functionality to facilitate interaction with the audience, a stopwatch feature for practice purposes, a feature to display key points on cue cards, and a memo function.

VII. Discussion

A. Cognitive Load of Real-Time Feedback

During presentations, real-time feedback can improve quality by allowing presenters to objectify their performance. However, this feedback sometimes leads to fatigue. Some participants found it difficult to assimilate real-time feedback while presenting. One noted feeling overwhelmed as EcoScript frequently updated vocal parameters like volume, speed, and pitch. Despite this, they recognized the benefit of understanding their vocal statistics. This supports Tanveer et al.'s research [13], which found intermittent feedback on vocal metrics more effective than continuous feedback. EcoScript's continuous updates and dynamic script tracking contributed to participant fatigue, highlighting the need to balance feedback frequency to prevent cognitive overload.

B. Comparative Analysis of Voice Features Stabilization

Log analysis revealed similar average vocal metrics for both tools, consistent with considerations of cognitive overload. It's hypothesized that presenters, focusing on content delivery, used feedback sporadically, resulting in similar average metrics. However, as shown in Figure 3, EcoScript had a lower standard deviation, indicating more consistent voice metrics. This aligns with participants' subjective ratings that EcoScript provided more ideal vocal metrics compared to PowerPoint.

Despite similar objective measures, subjective feedback remains crucial. Participants mentioned that EcoScript's feedback helped refine their presentation style and boosted confidence. This increase in self-awareness aligns with metacognitive theory, which relates to understanding and regulating one's cognitive processes [32]. EcoScript's real-time feedback enhances metacognitive awareness, facilitating better voice control and increasing presenter confidence.

C. The Need for Hands-Free Solutions

Our study validated EcoScript's automation features in real-time presentations. Features like automatic scrolling of cue cards and slide transitions aligned with the presentation script enable presenters to use natural gestures. Tailored automation enhances audience interaction and improves non-verbal communication, making presentations more natural and engaging for both presenters and audiences. These improvements are supported by research [11], [21], [23], laying the groundwork for an enhanced presentation experience and quality.

Participants praised EcoScript's automatic features but highlighted issues such as script misalignment or highlighting errors due to speech recognition inaccuracies. This underscores the need for mechanisms that allow swift correction of such errors to mitigate disruptions.

Currently, wireless control research is advancing in various ways [33], [34], and voice-driven screen management, as discovered in our study, is proving useful and could further evolve. For instance, this could include commands like "next, please" for slide transitions or emphasizing slides through commands like "this part is particularly important."

D. Limitations

Our study has the following limitations. First, the accuracy of speech recognition is not perfect, leading to potential errors in script tracking and causing confusion during presentations. Additionally, our experiments were conducted in a controlled, low-noise environment. To ensure robustness in real-world situations, noise-canceling techniques need to be added. Second, while we provided ideal voice feedback metrics to Korean speakers using a standard language, it is necessary to consider the feedback that reflects diverse vocal attributes, such as nationality and dialect. Lastly, the experiment focused primarily on enhancing presenters' skills, so evaluations were mostly conducted from the presenter's perspective, not from the audience's standpoint.

VIII. Conclusion

This study analyzed the difficulties faced by presenters and proposed a solution in the form of a software prototype called EcoScript. We identified key issues such as anxiety, lack of feedback, and difficulty in tracking content during presentations. To address these, we proposed EcoScript, which utilizes voice recognition technology to track scripts and provide real-time feedback on voice characteristics. Through user testing, EcoScript was found to enhance the convenience and overall quality of presentations using voice control. In conclusion, our research highlights the significance of leveraging voice recognition technology to provide real-time support to presenters during presentations, paving the way for advancements in presentation support tools.

Acknowledgments

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP)

grant funded by the Korea government(MSIT) (No. RS-2022-00155966, Artificial Intelligence Convergence Innovation Human Resources Development (Ewha Womans University)) and by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2024-2020-0-01460) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

REFERENCES

[1] Nowreyah A Al-Nouh, Muneera M Abdul-Kareem, and Hanan A Taqi. Efl college students' perceptions of the difficulties in oral presentation as a form of assessment. *International Journal of Higher Education*, 4(1):136–150, 2015.

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

[3] Claudio Baccarani and Angelo Bonfanti. Effective public speaking: a conceptual framework in the corporate-communication field. *Corporate Communications: An International Journal*, 20(3):375–390, 2015.

[4] Nick Morgan. The kinesthetic speaker. putting action into words. *Harvard business review*, 79(4):112–20, 2001.

[5] Timothy Clark and David Greatbatch. Audience perceptions of charismatic and non-charismatic oratory: The case of management gurus. *The Leadership Quarterly*, 22(1):22–32, 2011.

[6] Johan Roos. The benefits and limitations of leadership speeches in change initiatives. *Journal of Management Development*, 32(5):548–559, 2013.

[7] TVSS Swathi. *The importance of effective presentation for organizational success*. SSRN, 2015.

[8] J Elliot and Joyce LY Chong. Presentation anxiety: A challenge for some students and a pit of despair for others. *Available on: http://www. isana. org. au/files/20051017165939_PresentationAnxiety. pdf*, 2005.

[9] Abdullah Adnan Mohamed and Adelina Asmawi. Understanding engineering undergraduates' technical oral presentation: Challenges and perspectives. *International Journal of Language Education and Applied Linguistics*, 2018.

[10] Jan Schneider, Dirk Börner, Peter Van Rosmalen, and Marcus Specht. Presentation trainer, your public speaking multimodal coach. In *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pages 539–546, 2015.

[11] Mathieu Chollet, Torsten Wörtwein, Louis-Philippe Morency, Ari Shapiro, and Stefan 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*, pages 1143–1154, 2015.

[12] Mark Bubel, Ruiwen Jiang, Christine H Lee, Wen Shi, and Audrey Tse. Awareme: addressing fear of public speech through awareness. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*, pages 68–73, 2016.

[13] M Iftekhar Tanveer, Emy Lin, and Mohammed Hoque. 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*, pages 286–295, 2015.

[14] Tian Gan, Junnan Li, Yongkang Wong, and Mohan S Kankanhalli. A multi-sensor framework for personal presentation analytics. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 15(2):1–21, 2019.

[15] Clarence E Rash, Michael B Russo, Tomasz R Letowski, and Elmar T Schmeisser. Helmet-mounted displays: Sensation, perception and cognition issues. *US Army Aeromedical Research Laboratory*, 76, 2009.

[16] Andrew C Merkle, Michael Kleinberger, and O Manuel Uy. The effects of head-supported mass on the risk of neck injury in army personnel. *Johns Hopkins APL technical digest*, 26(1):75–83, 2005.

[17] Rebecca Hincks. Measures and perceptions of liveliness in student oral presentation speech: A proposal for an automatic feedback mechanism. *System*, 33(4):575–591, 2005.

[18] Xavier Ochoa, Federico Domínguez, Bruno Guamán, Ricardo Maya, Gabriel Falcones, and Jaime Castells. The rap system: Automatic feedback of oral presentation skills using multimodal analysis and low-cost sensors. In *Proceedings of the 8th international conference on learning analytics and knowledge*, pages 360–364, 2018.

[19] Xavier Ochoa. Multimodal systems for automated oral presentation feedback: A comparative analysis. In *The multimodal learning analytics handbook*, pages 53–78. Springer, 2022.

[20] Ha Trinh, Reza Asadi, Darren Edge, and T Bickmore. Robocop: A robotic coach for oral presentations. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(2):1–24, 2017.

[21] Reza Asadi, Ha Trinh, Harriet J Fell, and Timothy W Bickmore. Intelliprompter: speech-based dynamic note display interface for oral presentations. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 172–180, 2017.

[22] Luz Marina Quiroga, Martha E Crosby, and Marie K Iding. Reducing cognitive load. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, pages 9–pp. IEEE, 2004.

[23] Xavier Ochoa and Federico Dominguez. Controlled evaluation of a multimodal system to improve oral presentation skills in a real learning setting. *British Journal of Educational Technology*, 51(5):1615–1630, 2020.

[24] Everlyne Kimani, Ameneh Shamekhi, and Timothy Bickmore. Just breathe: Towards real-time intervention for public speaking anxiety. *Smart Health*, 19:100146, 2021.

[25] Alessandra Brandão, Hugo Nicolau, Shreya Tadas, and Vicki L Hanson. Slidepacer: A presentation delivery tool for instructors of deaf and hard of hearing students. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 25–32, 2016.

[26] React Developers. React documentation. https://react.dev/. [Online; accessed April 16, 2024].

[27] Sebastián Ramírez. Fastapi: Modern, fast (high-performance), web framework for building apis with python 3.7+ based on standard python type hints, 2024. Accessed: 2024-04-10.

[28] Joshua J Guyer, Leandre R Fabrigar, and Thomas I Vaughan-Johnston. Speech rate, intonation, and pitch: Investigating the bias and cue effects of vocal confidence on persuasion. *Personality and Social Psychology Bulletin*, 45(3):389–405, 2019.

[29] Pascal Belin, Bibi Boehme, and Phil McAleer. The sound of trustworthiness: Acoustic-based modulation of perceived voice personality. *PloS one*, 12(10):e0185651, 2017.

[30] Developer Name or Organization. Praat python library: Python library for automating praat software functionalities. https://github.com/praat/praat, 2024. Accessed: 2024-04-10.

[31] Google Cloud. Google cloud speech to text api documentation. https://cloud.google.com/speech-to-text, 2024. Accessed: 2024-04-10.

[32] Gregory Schraw and David Moshman. Metacognitive theories. *Educational psychology review*, 7:351–371, 1995.

[33] Salonee Powar, Shweta Kadam, Sonali Malage, and Priyanka Shingane. Automated digital presentation control using hand gesture technique. In *ITM Web of Conferences*, volume 44, page 03031. EDP Sciences, 2022.

[34] Felix Zhan. Hand gesture recognition with convolution neural networks. In *2019 IEEE 20th international conference on information reuse and integration for data science (IRI)*, pages 295–298. IEEE, 2019.