JOURNAL OF LEARNING ANALYTICS

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

Volume 11(3), 224–248. https://doi.org/10.18608/jla.2024.8411

OpenOPAF: An Open-Source Multimodal System for Automated Feedback for Oral Presentations

Xavier Ochoa[1], Heru Zhao[2]

Abstract

Providing automated feedback that facilitates the practice and acquisition of oral presentation skills has been one of the notable applications of multimodal learning analytics (MmLA). However, the closedness and general unavailability of existing systems have reduced their potential impact and benefits. This work introduces OpenOPAF, an open-source system designed to provide automated multimodal feedback for oral presentations. By leveraging analytics to assess body language, gaze direction, voice volume, articulation speed, filled pauses, and the use of text in visual aids, it provides real-time, actionable information to presenters. Evaluations conducted on OpenOPAF show that it performs similarly, both technically and pedagogically, to existing closed solutions. This system targets practitioners who wish to use it as-is to provide feedback to novice presenters, developers seeking to adapt it for other learning contexts, and researchers interested in experimenting with new feature extraction algorithms and report mechanisms and studying the acquisition of oral presentation skills. This initiative aims to foster a community-driven approach to democratize access to sophisticated analytics tools for oral presentation skill development.

Notes for Practice and Research

- Learning analytics practitioners can use the OpenOPAF system to offer additional practice-feedback cycles. This aids beginner presenters in acquiring foundational oral presentation skills without straining the resources of instructors or peers.
- Given its open-source nature, learning analytics technologists can customize the OpenOPAF system. This flexibility allows for support across various learner populations and adaptation to different learning environments.
- OpenOPAF is presented to the learning analytics research community as a common platform or testbed. It facilitates studies on multimodal feature extraction and the impact of automated feedback on the development of oral presentation skills.

Keywords

Open-source tool, communication skills, multimodal learning analytics.

Submitted: 08/03/2024 — **Accepted:** 09/10/2024 — **Published:** 13/11/2024

Corresponding author [1] Email: xavier.ochoa@nyu.edu Address: Steinhardt School of Culture, Education and Human Development, New York University, New York City, USA. ORCID iD: https://orcid.org/0000-0002-4371-7701

[2] Email: hz2296@nyu.edu Address: Steinhardt School of Culture, Education and Human Development, New York University, New York City, USA.

1. Introduction

Effective communication skills are crucial for the professional success of higher education students, yet most faculty members do not view developing these skills as their direct responsibility (Rios et al., 2020). This leads to a reliance on students gaining oral presentation skills through mere exposure, such as participating in course-related presentations (Thurneck, 2011). Although some educators incorporate these skills into their courses by requiring presentations, they often focus more on content than on communication technique, providing little substantive feedback on the latter. This approach results in inconsistent skill levels among graduates, with many not proficient in oral presentations (Donnell et al., 2011; Ochoa & Dominguez, 2020).

Mastery of oral presentation skills requires deliberate practice, which includes undertaking challenging tasks, receiving quantifiable performance feedback, and engaging in self-regulated learning (Ericsson et al., 1993; McGaghie et al., 2011). This cycle of practice and feedback, crucial for skill development, is frequently overlooked in teaching oral communication. Without structured feedback, students may continue making unnoticed mistakes, hindering their progress (Van Ginkel et al., 2015). To address this gap, learning analytics can provide essential performance metrics from practice sessions, offering students timely,

constructive feedback to foster improvement and mastery in their presentation skills (Chan, 2011).

Using multimodal learning analytics (MmLA) techniques, at least a dozen systems have been built in the last 15 years to provide some types of automated feedback for oral presentations in higher education settings (see Section 2.2 for a review). This general type of system will be referred to as oral presentation analysis and feedback (OPAF) in this work. Most of these systems have been tested with students and, in general, they seem to facilitate the acquisition of oral presentation skills (see Section 2.3 for evaluation results). However, two issues seem to indicate that there are two problems with the existing OPAFs: (1) there are 12 different systems that re-implement complex solutions for the same problems, and (2) even with positive results, and enough time to be polished into readily available products, they have not been adopted by the educational community.

The first issue, the need to build complex technological solutions from scratch, has been extensively discussed in the MmLA community as a major challenge in the field (Ochoa, 2022a; Yan et al., 2022). A recurring theme in regular MmLA workshops and symposia is the development of shared hardware and software platforms to capture and process relevant educational actions and interactions (Martinez-Maldonado et al., 2018; Worsley & Martinez-Maldonado, 2018). Responding to these calls, Li and colleagues (2024) proposed mBox, a bundle of sensors and algorithms designed to facilitate the multimodal study of collaborative learning settings. In another example, involving a higher-level architecture, Fernández-Nieto and colleagues (2024) shared an open-source platform, YarnSense, to provide an end-to-end pipeline for creating narrative-based visualizations based on multimodal data capture and processing. However, as will be clearly evident in this work's review of previous systems, there has been effectively no sharing of either hardware or software between the dozen OPAF systems previously developed and tested, with each using its own custom-built solution.

The second issue, the lack of adoption of these systems in real educational settings, is also common to other MmLA systems. Yan and colleagues (2022), in a systematic literature review, found that out of 96 MmLA systems, fewer than a quarter had been tested in ecologically valid environments (Technological Readiness Level 5—TRL5; Olechowski et al., 2015). They also found that none had reached TRL6 or higher. TRL6 is achieved when the system has demonstrated its usefulness in an operational environment. They conclude that although these systems can generate useful learning analytics information automatically and in real time, these benefits are not available to educational stakeholders during actual learning processes. The review conducted in this work (see Section 2.3) shows that this issue is even more prominent for OPAF systems, as only two out of 12 of these systems have been evaluated in an ecologically valid environment, with most of these systems still in the laboratory prototype phase.

Sharing hardware and software implementations among the MmLA research community seems to be the clear path to addressing the first issue. However, this approach does not directly address the second issue. Even well-designed frameworks, such as YarnSense (Fernández-Nieto et al., 2024), or general-purpose research tools, such as mBox (Li et al., 2024), are not as valuable to educational practitioners given the level of work still required to create ready-to-use tools from them. On the other hand, ready-to-use solutions that can be easily deployed in education, such as the OPAF system RAP (Ochoa et al., 2018), are usually not freely available to the community or are very difficult to change or adapt to other contexts. In an attempt to address these two issues simultaneously, this study introduces OpenOPAF, an open-source modular platform that replicates the functionalities and performance of existing systems on cost-effective, scalable hardware. By being openly shared, modular, and low cost, we expect that OpenOPAF will provide the following advantages over existing closed systems:

- It offers educational practitioners an accessible ready-to-use tool that can be deployed directly with students without requiring programming changes or licence fees.
- It provides educational technologists with an open platform that can be easily modified, contextualized, and integrated into other educational systems to facilitate adoption.
- It offers researchers a common, modular platform on which new extraction algorithms can be easily added and tested, new fusion and analysis pipelines can be implemented to improve the estimation of constructs of interest, and new feedback strategies can be explored.

This work will present the design and implementation of OpenOPAF, as well as a small evaluation study. The structure of this paper is organized as follows: Section 2 provides the necessary background for introducing OpenOPAF, including a brief description of what constitutes an OPAF system; a literature review of existing systems; and a summary of their documented impact on the acquisition of presentation skills. Section 3 details the guiding requirements and design decisions behind OpenOPAF, along with implementation specifics such as hardware components and software modules. Section 4 conducts an initial evaluation of OpenOPAF's performance in terms of technical accuracy, learning gains, and presenters' perception of the tool, to gain insights into its potential performance when compared with the state of the art. Finally, Section 5 outlines example usage scenarios for OpenOPAF and provides recommendations for its adoption across different groups in the learning analytics community.

ISSN 1929-7750 (online).

2. OPAF Systems

OPAF systems are the integration of software and hardware designed to capture multimodal information during oral presentations. They provide alerts, recommendations, or performance reports to presenters, aiding in the development or enhancement of their oral presentation skills. This definition includes a broad spectrum of systems, each with distinct educational objectives, implementations, and scales (Ochoa & Dominguez, 2020). Despite their diversity, all OPAFs follow a common internal process, akin to that of all MmLA systems (Ochoa, 2017). This section will review existing OPAFs and assess their effectiveness in improving oral presentation skills. For a more detailed overview and analysis of the state of the art of OPAF systems, the reader is invited to review the survey conducted by one of the authors (Ochoa, 2022b).

2.1 OPAF Operation

The process of any OPAF system begins with the recording of the presentation. This recording captures information for later analysis and feedback provision to the presenter. Typically, this information encompasses the visual and auditory details that human observers can readily perceive. Additionally, OPAFs may also record data usually imperceptible to humans, such as bio-signals (e.g., heart rate) or micro-movements (e.g., eye saccades). The recording is facilitated through one or more specialized sensors, which convert the observed information into digital formats for computational processing. Common sensor examples in OPAF systems include cameras (for visual and depth sensing) and microphones. An important consideration in sensor selection is their intrusiveness, or how comfortably the presenter can perform in the presence of these sensors. The outcome of this recording phase is a collection of different media types (video, audio, log files), each offering a distinct digital representation of the presentation.

The second step in the OPAF process involves extracting relevant modalities from the multimedia recording. These modalities represent the various methods through which the presenter conveys information to the audience (e.g., linguistics, paralinguistics, body language) or that reflect the presenter's mental state (e.g., arousal). It is possible to extract multiple modalities from the same medium; for instance, paralinguistic features like pitch and speech pace can be derived from audio recordings. Conversely, the same modality can be extracted from different media types; video recordings and joint-position logs, for example, can both be used to assess the presenter's posture. This step typically requires sophisticated signal processing algorithms (e.g., computer vision or speech processing) and powerful computing resources to achieve real-time (or near-real-time) processing.

Following the extraction of modalities, the next phase is to analyze these modalities to identify the presenter's actions and assess the presentation's quality. Initially, low-level features are extracted from the modalities. For example, variations in pitch over time could represent such a feature for the pitch modality, whereas posture could be categorized as open or closed. Extracting these features is generally less complex than the initial modality extraction from the recording. Optionally, these low-level features can be fused to form more complex, robust features. For instance, combining features from body micro-movements (e.g., trembling) with stuttering patterns from the paralinguistic modality could help detect presenter nervousness. Lastly, these low- and mid-level features are analyzed to identify behavioural indicators that assess mastery over various oral presentation skills, such as using gaze direction analysis to evaluate eye contact with the audience.

The final step in the OPAF process is generating feedback for the presenter. Although this might seem straightforward, it is arguably the most complex design aspect due to the myriad variables that influence the end-user's perception and the system's educational efficacy. The primary feedback variable is its content, which not only relates to the behavioural indicators identified previously but may also include improvement recommendations, comparisons with past performances, or benchmarks. Feedback presentation is another critical variable, encompassing the modalities (visual, aural, haptic) used for feedback delivery and the timing of feedback (during or after the presentation). Both the timing and modality of feedback delivery significantly affect the system's intrusiveness.

While the described process is a generalization, specific systems may vary from this model. For instance, depth sensors like the Microsoft Kinect might perform modality extraction (e.g., presenter's posture) directly in the recording phase, often capturing skeletal joint positions rather than raw depth data. Although deviations from this streamlined process exist, the simplified overview aids in discussing existing OPAFs and introducing OpenOPAF in subsequent sections.

2.2 Existing Systems

The first comprehensive OPAF system documented in the literature, "Presentation Sensei," was developed by Kurihara and colleagues (2007) in Japan to assist higher education students in rehearsing their presentations. This pioneering system used computer vision and speech processing algorithms to facilitate the development of presentation skills. Six years later, Batrinca and colleagues (2013) introduced a significantly more sophisticated system named "Cicero," which was designed to automatically evaluate oral presentations. Beyond employing advanced sensors, it featured a virtual interactive audience that mimicked the visual and auditory feedback of real audiences.

ISSN 1929-7750 (online).

Between 2015 and 2016, reports surfaced about several independent yet similar systems. Dermody and Sutherland (2015) and Schneider and colleagues (2015) developed systems leveraging the Microsoft Kinect Sensor to offer feedback to presenters during rehearsal sessions. In contrast, "Logue," created by Damian and colleagues (2015), used a wearable sensor (Google Glass) to capture egocentric perspectives from both the presenter and audience members, providing feedback to enhance speaker performance. Concurrently, "PresentMate" (Lui et al., 2015) employed mobile phone sensors to record presenters' movements and speech, offering feedback across practice, performance, and debriefing stages. Additionally, Nguyen and colleagues (2015) introduced a system aimed at automatically assessing presentations, while Tanveer and colleagues (2015) proposed a system to improve speaker performance during live presentations. Finally, Gan and colleagues (2015) generated a dataset and system to evaluate oral presentations in Singapore.

In more recent years, innovative systems have emerged, leveraging advancements in sensor and artificial intelligence technologies. RAP (Ochoa et al., 2018) substitutes the Microsoft Kinect with a simple webcam and deep learning–based computer vision algorithms to provide feedback for early-year student presenters during rehearsals. RoboCOP (Trinh et al., 2017) features a humanoid robot head offering verbal and non-verbal feedback, emulating advice from human experts. Moreover, an updated version of the Presentation Trainer (Schneider et al., 2019) employs virtual reality (VR) to facilitate feedback during rehearsal sessions.

As shown in Table 1, these systems were developed almost concurrently across Europe, Asia, and the Americas. While they acknowledge early contributions in their references, particularly those of Kurihara and colleagues (2007) and Batrinca and colleagues (2013), there appears to be no collaboration among the various research groups, as they do not cite each other. None of the reviewed systems are currently openly or readily available to interested parties outside their original research groups. Based on the authors' professional connections with many of these research teams, it is known that most of the original researchers have moved on to other groups and are now working on different projects.

Table 1. List of OPAF systems reported in the literature.

Name	Citations	Year	Focus	Countries
Presentation Sensei	Kurihara et al.	2007	Training	Japan
Cicero	Batrinca et al.	2013	Live Event	Italy, USA
Logue	Damian et al.	2015	Live Event	Germany, Belgium
NA	Dermody and Sutherland	2015	Training	Ireland
NUSMSP	Gan et al.	2015	Live Event	Singapore
PresentMate	Lui et al.	2015	Training	China
NA	Nguyen et al.	2015	Training	Netherlands
Presentation Trainer	Schneider et al.	2015	Training	Netherlands
Rhema	Tanveer et al.	2015	Live Event	USA
RoboCOP	Trinh et al.	2017	Training	USA
RAP	Ochoa et al.	2018	Training	Ecuador
Presentation Trainer VR	Schneider et al.	2019	Training	Netherlands

2.3 Evaluation

OPAF systems have been evaluated through various methods. These evaluations can focus on the technical accuracy of modalities extraction, presenters' perceptions of the system, improvements in presenters' abilities with repeated system use, or the impact of the system on students' presentation skills as assessed by experts (Ochoa & Dominguez, 2020). For the purpose of assessing OPAFs' effectiveness in enhancing real-world oral presentation skills, this subsection will specifically discuss evaluations based on expert judgments of presentation quality.

To date, only four studies have attempted to determine the actual effectiveness of OPAFs, with varying levels of success. Tanveer and colleagues (2015) conducted the first evaluation of their system, Rhema. Their research involved a laboratory-controlled experiment with 30 participants who were randomly assigned to control and intervention groups. The presentations were recorded and rated by 10 judges using a basic rubric. These judges were laypeople with no expertise in oral presentations. Perhaps due to these limitations, the study found no statistically significant difference in presentations with or without the system. Schneider and colleagues (2016) conducted an "in-the-wild" study and reported a statistically significant positive impact on peer-evaluated scores after using the Presentation Trainer system. However, this study was not controlled for external variables and involved only nine students. Trihn and colleagues (2017) assessed learning transfer after system use in a laboratory setting with 12 participants and 12 judges, who had varying levels of presentation experience. The study observed statistically significant improvements in only a few aspects of presentations. The most comprehensive study, conducted by Ochoa and Dominguez (2020), was a controlled experiment involving 180 student presenters, 85 of whom used the system.

ISSN 1929-7750 (online).

They discovered that the RAP system had a statistically significant positive effect on student performance compared to those who did not use the system, as evaluated by experts.

The limited number of evaluations of OPAF systems yields generally positive, albeit nuanced, results. Most evaluations, especially those in controlled settings, indicate an overall improvement in oral presentation skills for users of OPAFs. However, not all aspects of oral presentation skills are equally trainable in a few sessions. For instance, maintaining eye contact with the audience appears to be an easily acquired skill with OPAFs, while reducing the use of filled pauses may require more practice and possibly different types of feedback than those currently employed (Ochoa & Dominguez, 2020).

An important aspect to consider is that these evaluations have been conducted primarily in laboratory settings, with only the Presentation Trainer and RAP being evaluated in ecologically valid environments. Additionally, to the best of the authors' knowledge, based on professional connections with many of the original authors, none of these systems are currently being used regularly in any real educational context. This situation aligns with the findings of Yan and colleagues (2022) regarding the low level of technical maturity of MmLA tools.

2.4 Conclusions

The previous review of existing systems demonstrates that creating systems for automated feedback on oral presentations has been both technically feasible and pedagogically effective. Numerous groups around the world have developed and tested independent, yet similar, OPAF systems with this goal in mind. However, two notable points also emerge from this review: (1) none of the reviewed OPAF systems have been built using technology from previous systems, and (2) none of the OPAF systems have been adopted for use in real educational contexts. These issues, already identified by the MmLA community as a whole (Martinez-Maldonado et al., 2018; Worsley & Martinez-Maldonado, 2018; Yan et al., 2022), unsurprisingly also apply to OPAF systems.

As mentioned in the introduction, OpenOPAF seeks to address these two issues by providing a modular, low-cost, open-source implementation of OPAF system functionalities. By being affordable and freely accessible, without requiring permission or collaboration with the original creators, OpenOPAF could facilitate the adoption of this technology by any interested party (Williams van Rooij, 2011). Additionally, open hardware and software have been successfully shared in many fields (Clegg et al., 2022; Coleman & Salter, 2023) to promote collaboration between research groups and accelerate progress. The next section will introduce OpenOPAF, a project designed to expand the use of OPAFs in real educational contexts and facilitate the design and implementation of more advanced systems.

3. OpenOPAF

OpenOPAF aims to offer researchers and practitioners an accessible platform for deploying, customizing, assessing, and enhancing automated systems for analyzing and providing feedback on oral presentations. To clarify how OpenOPAF operates, the rationale behind its design choices, and the potential for its application and expansion, this section will detail its foundational design objectives, the types of data it analyzes, the hardware and software infrastructure, its intended operational context, and the terms under which it is made available to the community.

The review of existing OPAF evaluations (Section 2.3) found that, in general, these systems are both useful and effective. As discussed in the previous section, the main issues with the current generation of OPAFs are their limited availability to third parties and their lack of adoption in real educational contexts. In this context, creating an unproven novel implementation of an OPAF would not be productive. Instead, we chose to reverse-engineer (Motavalli, 1998) existing systems, particularly the most recent and well evaluated system, RAP (Ochoa et al., 2018). The goal was to replicate the functionality and features of these systems to maintain their effectiveness while avoiding potential copyright violations and enhancing affordability. Consequently, OpenOPAF does not introduce significant innovations compared to existing systems and is not the result of a complete, original design cycle. However, within the constraints of the reverse-engineering process, several design decisions were made to create the final OpenOPAF system. These decisions are detailed in the following subsections.

3.1 Design Decisions

Even though OpenOPAF is a reverse-engineered version of existing OPAFs, there were still design decisions to be made, as each of these systems had its own particular set of choices. The design decisions for OpenOPAF, while constrained by this original framework, were selected to align with the most common and successful approaches identified in the systems reviewed in Section 2.2. These prior designs, combined with the need for affordability and modularity, guided the specific design choices detailed below.

1. **Focus on non-verbal skills:** All reviewed OPAFs provide feedback designed to train basic non-verbal skills, such as making eye contact with the audience and speaking loudly enough (see Section 3.2). Following this approach, OpenOPAF will also focus on training these basic skills to support beginner presenters. Although the effectiveness of OPAFs in

training some of these skills has not yet been demonstrated (Ochoa & Dominguez, 2020), we have chosen to include them all to provide future researchers with baselines for comparing new solutions. To provide feedback on basic non-verbal skills, OpenOPAF will capture and analyze the modalities needed to measure and assess these skills (see Section 3.3), replicating the capabilities of existing OPAFs in analyzing body language, non-verbal aural communication, and visual aids.

2. **Focus on practice, not evaluation:** OPAFs that emphasized presentation practice, rather than just-in-time feedback or evaluation, have been shown to significantly improve the acquisition of presentation skills (Ochoa & Dominguez, 2020). Following this trend, OpenOPAF primarily provides formative feedback during practice sessions, rather than summative feedback during or after actual presentations.

3. **Non-intrusive equipment:** As discussed in Section 2.2, previous OPAFs vary widely in the types of measuring equipment used to capture the presenter's behaviour. Given its focus on adoption in real educational settings, OpenOPAF uses hardware that avoids the need for wearables or disruptive devices, thereby maintaining a realistic rehearsal environment conducive to natural presentation behaviours. In this respect, OpenOPAF is more similar to the RAP system (Ochoa et al., 2018) than to the headset-requiring Presentation Trainer VR (Schneider et al., 2019).

4. **In-the-wild scalability:** Designed for practical use beyond laboratory settings, OpenOPAF is compatible with resources typically available in higher education institutions in high- and middle-income countries. It operates with sensors that adapt to standard environments without requiring special arrangements or controlled conditions. For example, it can capture and extract speech features such as volume or articulation rate in the presence of typical classroom noise and record body language under various classroom lighting conditions. The system is designed to be low-cost (less than 500 USD) to ensure that it is affordable for adoption in the mentioned types of institutions.

5. **Modularity:** A modular approach enables researchers and developers to select and combine different components and features based on their specific requirements, ensuring that the system remains flexible and adaptable to various use cases. Moreover, it provides an extensible framework to which new sensors, modalities, feature extraction algorithms, score calculations, and reporting elements can be easily added or modified with minimal effort.

6. **Privacy:** In previous OPAF systems, data privacy was not a primary concern, as they were mainly intended as laboratory prototypes for research. However, OpenOPAF is designed for use in real educational environments. As such, it ensures privacy by controlling data access through a unique username and password, and it allows presenters to maintain data ownership by offering the option to permanently delete their data from the device.

These design decisions, mostly based on existing OPAF designs, always include a trade-off, particularly between covering multiple modalities and using non-intrusive equipment, which will be further discussed in subsequent sections.

3.2 Skills to Train

According to design decision 1, the main pedagogical focus of OpenOPAF will be training non-verbal oral presentation skills. These skills, which are defined by the use of body language, facial expressions, gestures, eye contact, posture, paralinguistic features, and other physical cues to effectively convey a message and engage an audience during a presentation, are often cited in the literature as key characteristics of an accomplished oral communicator (Subapriya, 2009; Bull & Frederikson, 2019; Kilag et al., 2023). As important as they are, these types of skills have been found to be lacking in novice presenters (Castañer et al., 2013; Domínguez et al., 2023). It is not surprising, then, that OPAFs, with a general focus on training novice presenters, concentrate on measuring and providing feedback mainly for these types of skills.

Table 2 shows the non-verbal skills for which existing OPAF systems provide feedback. These skills are as follows:

- **Looking at the audience** involves making regular and intentional eye contact with the people in the room. This skill helps establish a connection with the audience, making them feel engaged and valued (Kilag et al., 2023). By scanning the room and briefly making eye contact with different individuals or sections of the audience, the presenter can convey confidence, sincerity, and attentiveness. This also helps in gauging the audience's reactions, allowing the presenter to adjust their delivery in real time. Novice presenters frequently look at the slides or away, rather than toward the audience (Domínguez et al., 2023).

- **Maintaining an open posture and gestures** involves standing or sitting with the body facing the audience directly, shoulders back and relaxed, and feet shoulder-width apart for a stable, confident stance. Presenters should use open hand gestures with visible palms, making broad, inclusive movements to engage the audience. Presenters should avoid crossing their arms or fidgeting, as these can signal defensiveness or nervousness. This approach helps convey confidence,

ISSN 1929-7750 (online).

Table 2. Skills trained by existing OPAF systems. Look = Looking at audience, PosGes = Maintaining an open posture and gestures, Vol = Maintaining good voice volume, Cad = Maintaining an adequate speech cadence, FilPau = Avoiding filled pauses, TLen = Avoiding too much text in the visual aids, FSize = Using a visible font size in the visual aids.

Name	Modalities						
	Body Language		Aural			Visual	
	Look	PosGes	Vol	Cad	FiPau	TLen	FSize
Presentation Sensei	X			X	X		
Cicero	X	X	X		X		
Logue		X		X			
NA—Dermondy and colleagues	X	X		X	X		
NUSMSP	X	X	X	X	X		
PresentMate			X				
NA—Nguyen and colleagues	X	X					
Presentation Trainer	X	X	X		X		
Rhema			X	X			
RoboCOP	X				X	X	
RAP	X	X	X		X	X	X
Presentation Trainer VR	X	X	X		X		

approachability, and honesty during communication (Van Ginkel, 2019). Novice presenters often exhibit a defensive or closed posture, such as crossing arms or placing hands in pockets (Domínguez et al., 2023).

- **Maintaining good voice volume** means speaking loudly enough to be heard clearly by everyone in the room without shouting. The voice should be projected confidently, filling the space, while remaining natural and comfortable enough to be sustained throughout the presentation. This helps ensure that the audience clearly hears the communicated message (Kilag et al., 2023). Some novice presenters speak too softly to be clearly audible to the audience (Domínguez et al., 2023).
- **Maintaining an adequate speech cadence** involves pacing speech at a rhythm (usually measured in words per minute) that is clear and engaging, avoiding both rapid, rushed delivery and overly slow, monotonous speech. This balanced cadence helps keep the audience's attention, making the presentation more dynamic and easier to follow (Dowhower, 1991). Novice presenters often struggle to control their speaking pace, usually speaking too fast (Domínguez et al., 2023).
- **Avoiding filled pauses** means minimizing the use of filler words like "um," "uh," and "like" during speech. The speaker should allow for natural pauses instead of filling them with unnecessary sounds. This creates a smoother, more professional delivery (Kilag et al., 2023).
- **Avoiding too much text in the visual aids** involves designing slides or other visual materials with minimal text to ensure clarity and focus. This skill helps prevent overwhelming the audience with information and encourages the presenter to speak directly to the key points rather than reading from the slides. By using concise bullet points, keywords, or visuals instead of lengthy paragraphs, the presenter can maintain the audience's attention, making the content more engaging and easier to absorb (Alley & Robertshaw, 2004).
- **Using a visible font size in visual aids** involves selecting a typeface with a size large enough to be easily read by all audience members, even those seated at the back of the room. This skill ensures that the information on the slides or other visual materials is accessible to everyone, preventing frustration or disengagement caused by illegible text (Alley & Robertshaw, 2004).

While these are not all the non-verbal skills needed for a successful oral presentation, they are all important (Kilag et al., 2023; Van Ginkel, 2019; Alley & Robertshaw, 2004), and they are the ones for which we can currently generate automated feedback. OpenOPAF will be designed to address this list of seven skills previously implemented by existing OPAFs.

3.3 Covered Modalities

To provide feedback that can help presenters practise and improve the skills discussed in the previous section (Section 3.2), OpenOPAF will extract and analyze modalities from audio, video, and visual aids.

From the audio, OpenOPAF will extract three aural modalities: (1) Voice Volume captures the average amplitude of the presenter's voice. This data will be used to provide feedback on the skill of "Maintaining a good voice volume." (2)

Articulation Rate corresponds to the speed at which the presenter speaks, measured in words per minute. This rate will be used to provide feedback on the skill of "Maintaining an adequate speech cadence." (3) Filled Pauses detects and quantifies non-lexical utterances, providing feedback on the skill of "Avoiding filled pauses."

From the video, OpenOPAF will extract two body language modalities: (1) Gaze Direction measures the direction of the presenter's gaze or head orientation. This data will be used to provide feedback on the skill of "Looking at the audience." (2) Body Posture assesses the presenter's stance by analyzing the positions of key articulation joints. This will be used to provide feedback on the skill of "Maintaining an open posture and gestures."

Finally, from the visual aids, OpenOPAF will extract two modalities: (1) Text Length analyzes the amount of text on each slide. This modality will provide feedback on the skill of "Avoiding excessive text in visual aids." (2) Font Size evaluates the typeface sizes used on each slide, providing feedback on the skill of "Using a visible font size in visual aids."

The capture and extraction of these modalities will be implemented in the initial version of OpenOPAF. Leveraging the system's modularity, additional modalities can be integrated to target other presentation skills.

3.4 General System Design

In accordance with design decision 5, the OpenOPAF system is designed to be highly modular, with the architecture organized into separate, interchangeable components, each handling a specific aspect of the system's functionality. This modularity enhances the system's maintainability, testability, and extensibility. The general architecture of the system is depicted in Figure 1. Each of its layers is described in the following subsections.

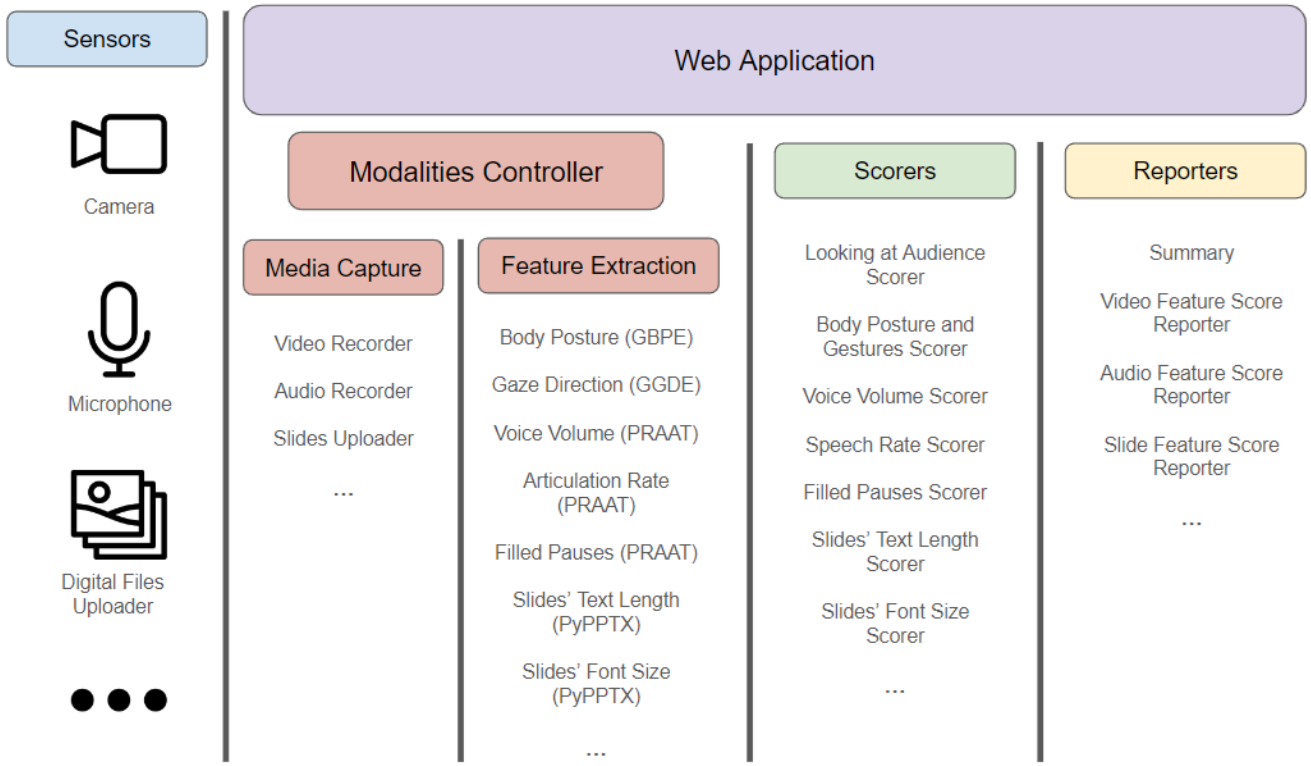


Figure 1. Overview of OpenOPAF's design, showing the different component types (coloured boxes).

3.4.1 Sensors and Hardware

The hardware for constructing OpenOPAF was chosen to balance design decisions 3 (non-intrusive equipment), 4 (in-the-wild scalability), and 6 (privacy). This design provides the necessary performance for multimodal analysis while processing data streams locally and minimizing costs. The hardware uses off-the-shelf components capable of analyzing audio and video signals in real time during presentations. The core components cost less than 300 USD, with options to include additional elements based on application needs. Specifications and source files for 3D-printable parts are available in the hardware section of the companion repository to this paper[1].

[1] Hardware specifications: https://github.com/xaoch/OpenOPAF/blob/main/hardware/hardware.md

ISSN 1929-7750 (online).

The core OpenOPAF device consists of four elements: a processing unit, a microphone, a camera, and a case (see Figure 2). The NVIDIA Jetson Nano Developer Kit 4GB[2] was chosen as the processing unit. This low-cost microcomputer features an integrated Tegra-class GPU, with prices ranging between 100 and 200 USD. The Jetson Nano was selected over the commonly used Raspberry Pi microcomputer due to its capacity for executing real-time deep neural network algorithms for feature extraction.

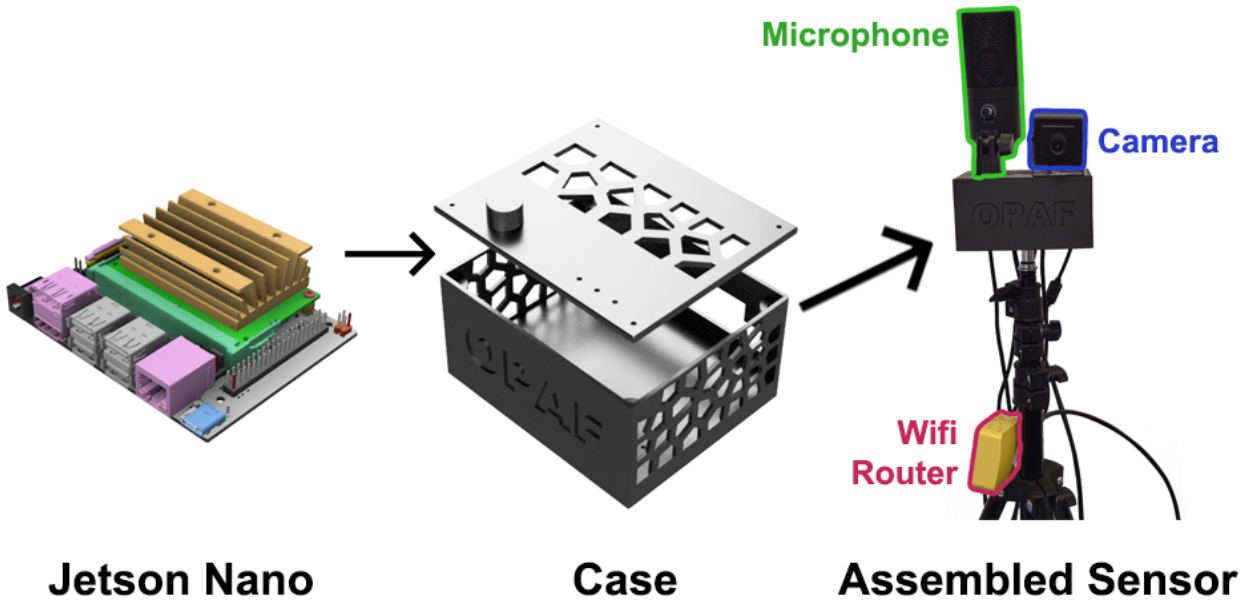


Figure 2. OpenOPAF core hardware components.

For audio capture, any high-quality microphone, such as a prosumer microphone designed for podcasters[3], is suitable. The primary requirement is USB connectivity to the Jetson Nano, as it lacks an audio jack. These microphones typically cost between 30 and 50 USD. For video capture, any camera capable of recording at 15 frames per second with a resolution above 800×600 will suffice. Consumer-level web cameras from brands like Logitech or ELP[4] meet these criteria and cost between 30 and 70 USD. Both the microphone and the camera should offer mounting options for integration with the device. To comply with design decision 5 (modularity), the hardware subsystems allow for the inclusion of other types of sensors (e.g., the Empatica wristband sensor (McCarthy et al., 2016) to measure stress levels), which can be connected via USB, Wi-Fi, or Bluetooth.

The device's case can be 3D printed to house the processing unit and mount the camera and microphone. Key considerations for the case design include adequate ventilation for the Jetson Nano's heat dissipation fins, openings for all of the Nano's ports, camera and microphone mounting points, and options for tripod or wall mounting. Because the source files for the case design are also shared, the case can be modified to include attachment points for other types of sensors.

OpenOPAF may require additional equipment for deployment in specific environments, with details available in the companion repository. Necessary setups include a computer and projector or screen: one to display the system interface and a virtual audience in front of the student, and another to present the speaker's slides behind them.

Lacking built-in Wi-Fi, the Jetson Nano requires a mini Wi-Fi router for wireless connectivity and internet access, which is useful for email report delivery. To ensure data security and privacy, the Wi-Fi network can be isolated, with offline report access via removable storage. The OpenOPAF device should be positioned centrally in front of the speaker, necessitating a compact and slim tripod or wall/ceiling mount to avoid obstructing the presenter's view. Additionally, the Jetson Nano's GPU requires a compatible 20-watt power source, rather than a standard USB charger.

3.4.2 Software Architecture

The NVIDIA Jetson Nano operates on an Ubuntu Linux-based operating system (OS) called NVIDIA JetPack. Built on this OS is a custom Python Flask Web Application (represented by the purple component in Figure 2), which provides a

[2]NVIDIA Jetson Nano: https://developer.nvidia.com/buy-jetson?product=jetson_nano
[3]Example of the microphone used in the standard implementation: https://www.amazon.com/gp/product/B06XCKGLTP
[4]Example of the camera used in the standard implementation: https://www.amazon.com/gp/product/B06ZXW6QBV

web-based control interface. This interface manages device operation, communicates with sensors (camera and microphone) for presentation recording, extracts features from the recorded modalities, calculates and generates feedback reports, and tracks user progress over time. The source code of the Python application, along with instructions for installing the operating system and configuring the Jetson Nano, can be found in the OpenOPAF project repository[5].

Within this application, a control interface allows users to navigate the system independently without requiring technical assistance. To use the control interface (shown in Figure 3), users must first log in or register with a new username and password. The interface provides options to initiate a new practice session, review past reports, or view a dashboard summarizing their progress. For new practice sessions, users can set the session duration and choose whether to include slide presentations for analysis. Before beginning a session, the system offers a camera preview to ensure full-body visibility, allowing users to adjust their positioning as needed. Subsequently, a virtual audience and timer are displayed, prompting users to commence their presentation.

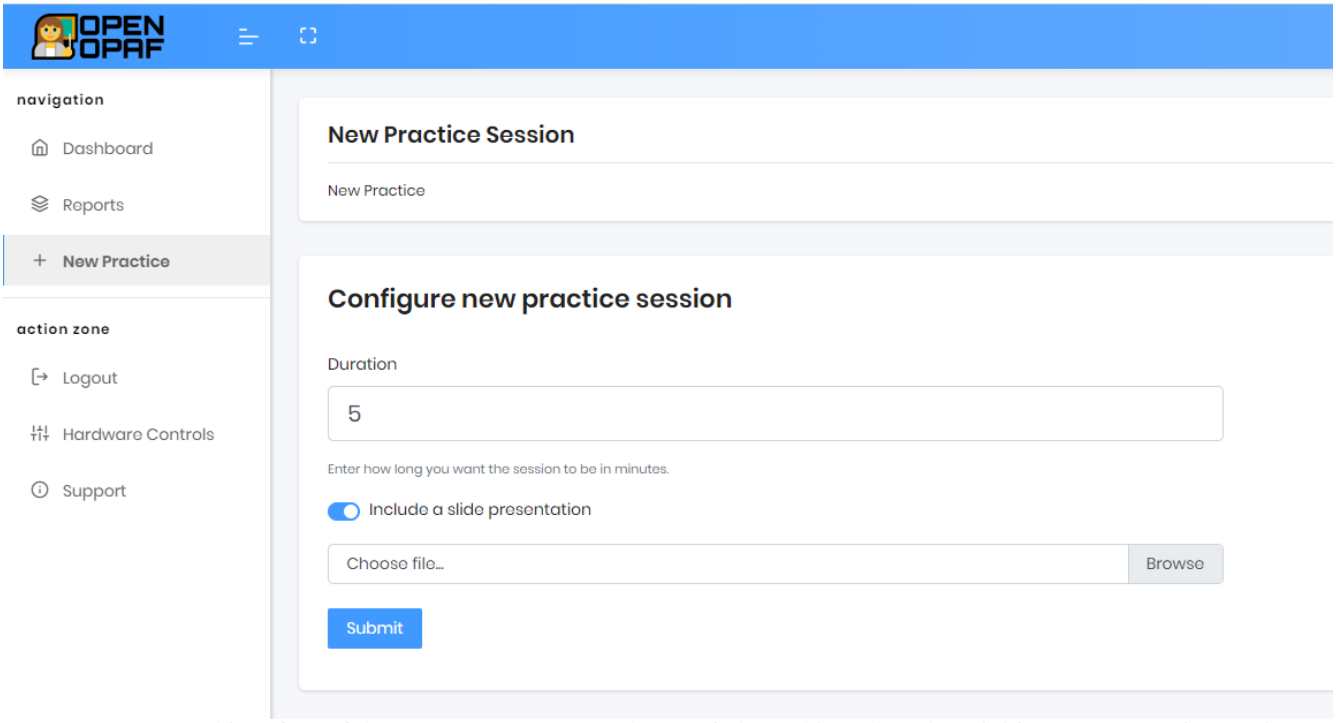


Figure 3. Control interface of the OpenOPAF system. The user is logged in and ready to initiate a new practice session.

To comply with design decision 5 (modularity), the internal design of the OpenOPAF application is divided into four compartmentalized sections, each containing a variable number of interchangeable components. The first section, Media Capture, manages the start and stop functions for recording different types of media obtained from the sensors and provides the digital files used later in feature extraction. Currently, three components are implemented: a video recorder, an audio recorder, and a slides uploader. The video recorder captures video at 640×480 resolution at two frames per second using the OpenCV Python library[6]. The audio recorder captures mono audio at 44,100 Hz, 16-bit resolution using the SoundDevice Python library[7]. The slides uploader allows Microsoft PowerPoint slides in "pptx" format to be loaded into the system. If a new or different type of sensor is connected, a new Media Capture component should be implemented to support the Media Capture interface.

While the modality streams are being recorded, the Modalities Controller activates different Feature Extraction components. Features for each modality are extracted in real time during the presentation recording. The currently implemented Feature Extraction components are as follows:

- **Video Feature Extraction:** Currently, two components have been implemented: a Geometric Body Posture Extractor (GBPE) and a Geometric Gaze Direction Extractor (GGDE). First, each video frame is analyzed using the holistic landmark detection model from the MediaPipe library (Chunduru et al., 2021). This neural model extracts the 3D

[5]OpenOPAF Github repository: https://github.com/xaoch/OpenOPAF/tree/main
[6]OpenCV Python library: https://github.com/opencv/opencv-python
[7]SoundDevice Python library: https://github.com/spatialaudio/python-sounddevice

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License (CC BY 4.0)

positions of 33 body landmarks and 478 face landmarks. This extraction method has been shown to be highly accurate in identifying postures (97%), even in complex human poses (e.g., yoga practice) in medium-quality images similar to those captured by OpenOPAF (Debalaxmi et al., 2024). These landmarks are used to estimate both body posture and gaze direction. The GBPE relies on a geometric detector that identifies erroneous postures, such as when the hands are at the sides of the body (e.g., hands in pockets), too close to the face, or in front of the torso (indicating a closed posture), or when one shoulder is misaligned with the other (indicating body orientation away from the audience). The GGDE estimates gaze direction through a 3D projection approach that uses the positions of the eyes, nose, ears, and mouth. These landmarks help calculate face rotation around the z-axis (yaw) and y-axis (pitch), determining whether the presenter is focused on the audience or looking elsewhere. This extraction process matches the video recording speed of 15 frames per second. In line with design decision 5 (modularity), different video extractors based on neural network models (Chong et al., 2020), for example, could be used to translate the original frame or the landmarks into postures or gaze direction, or any new modality.

- **Audio Feature Extraction:** Every 5 seconds of audio recording is analyzed using the Parselmouth Python library[8], an implementation of the PRAAT audio library (Boersma & Van Heuven, 2001). Three extractors are currently implemented using different PRAAT scripts (De Jong et al., 2021) that estimate speech features by excluding non-speech periods. The first extractor determines the Voice Volume by the average power of the speech signal during these periods. The second extractor measures Articulation Rate by detecting syllabic components, counting them over 5 seconds, and converting this to words per minute (assuming an average of 1.66 syllables per word) (De Jong & Wempe, 2009). The third extractor, Filled Pauses, identifies filler words by locating prolonged syllables with minimal spectral contrast (De Jong et al., 2021). These features are calculated at 5-second intervals. The accuracy of the speech volume measurement is reliable, as it depends solely on the sensor's precision. The estimation of speech rate has been found to highly correlate (Pearson's r coefficient = 0.82) with manual measurement in English speech (De Jong & Wempe, 2009) under low-noise conditions. The detection of filled pauses in English has an accuracy of 84% in low-noise environments (De Jong et al., 2021). However, there is no literature measuring the accuracy of these extractions in high-noise environments. New extractors could be added to obtain different audio features, such as pitch variation (Deshmukh et al., 2005).

- **Slides Feature Extraction:** At the start of the session, the digital Microsoft PowerPoint file containing the slides is uploaded by the Slides Uploader component. The pptx Python library[9] analyzes the content of each slide, extracting all text and measuring two properties. This information is passed to the two implemented feature extractors: Slides' Text Length and Slides' Font Size. Default template sizes are assumed to be 18 points if no specific size information is provided. Since both of these measurements are direct counts from a digital file, the values provided are always accurate.

The modular design of the extractor components allows for the incorporation of new techniques or algorithms to enhance feature extraction accuracy, interpret features more effectively, or introduce new modalities in future versions.

Upon completion of the recording session, the features extracted from the different modalities are used to score the presenter's performance. This step translates the raw modality features into an assessment of the different skills trained by the system (see Section 3.2). Currently, each of the six skills targeted by OpenOPAF has a corresponding scorer module (see Figure 1). These scorers contain adjustable thresholds that determine the general level of performance in a given skill. For example, a presenter is currently considered to have mastered the skill of looking at the audience if they maintain their gaze on the audience for at least 85% of the time. Since these thresholds depend on culture and learning objectives (Van Ginkel et al., 2017), their values are meant to be tailored according to the expectations of each educational context.

Finally, after the scores have been calculated, they are compiled, along with the raw recordings and extracted features, to create a feedback report. These feedback reports are assembled in less than 15 seconds to allow for immediate feedback after practice, as guided by design decision 2 (focus on practice). The report aims to give presenters a chance to reflect on their presentation skills by observing and listening to themselves, alongside the system's performance estimations. The feedback report includes four types of components:

- **Summary:** This component provides an overview of the presentation alongside improvement recommendations based on the extracted feature values across different modalities (see Figure 4). A textual description of performance is paired with a visual summary, indicating performance levels with configurable thresholds, currently set for novice users.

- **Video-Based Reporter:** The report offers an interactive timeline synchronized with the video recording for video-based modalities (see Figure 5). Video frames are aggregated into reviewable segments based on statistical mode (currently

[8]Parselmouth Python library: https://parselmouth.readthedocs.io/en/stable/
[9]pptx Python library: https://github.com/scanny/python-pptx

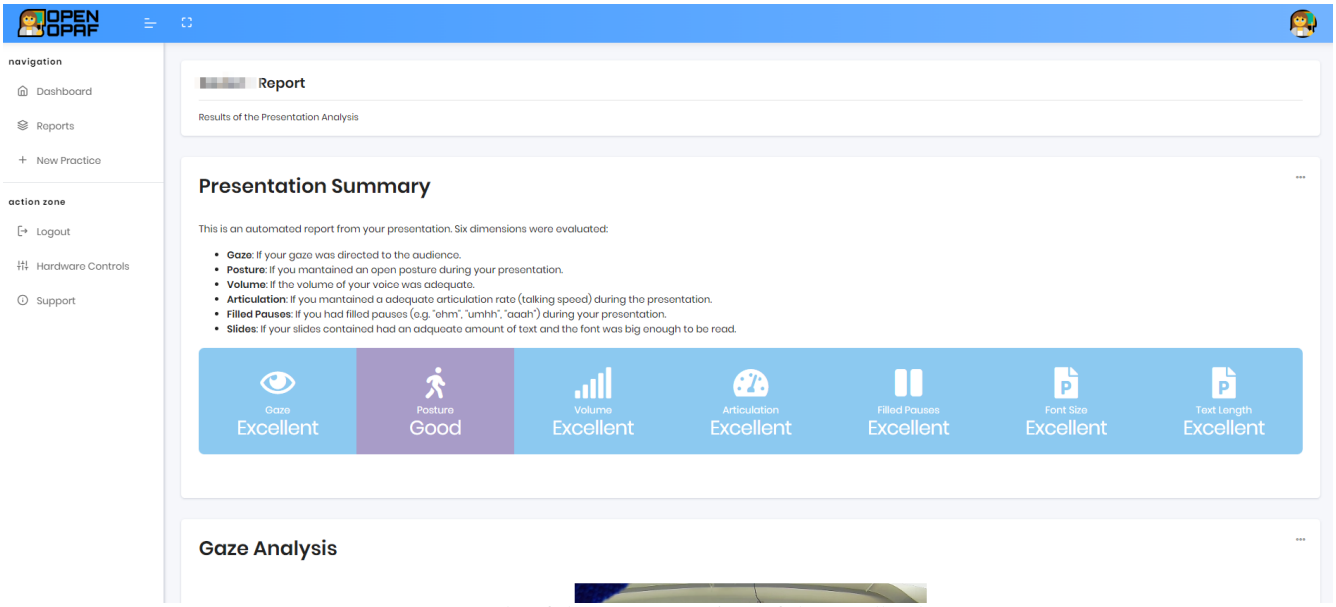


Figure 4. Example of the summary view of the Feedback Report.

set at 5-second intervals, but adjustable). The timeline's colour coding indicates posture correctness, with light blue for correct, open postures and pink for non-recommended postures.

- **Audio-Based Reporter:** Audio-based modalities are represented by an interactive waveform (bottom of Figure 6) and a line chart displaying feature values at each 5-second interval (top of Figure 6). Users can select waveform segments to listen to their speech during those intervals. Background colours in the waveform indicate value correctness, with light blue for recommended values, purple for borderline, and pink for non-recommended values, with customizable thresholds.

- **Slides-Based Reporter:** For slides-based modalities, the report features a timeline displaying extracted feature values for each slide (bottom of Figure 7), linked to slide images (top of Figure 7). Clicking on the timeline updates the slide being displayed. Timeline colours reflect the evaluation of the slide's text (light blue: correct, pink: not recommended).

Similar to the feature extraction modules, the report's format and content can be adapted to fit the learning context of the system's application.

3.5 Operation Environment

OpenOPAF is designed as a portable solution that can be easily set up and removed in medium- to small-sized rooms, such as a small classroom, a meeting room, or a study room. The main constraint on the minimum size of the space is determined by the distance required for the camera to capture the full body of the presenter. For a standard webcam with normal lenses, this distance is approximately 3 metres. While wide-angle lenses can reduce this distance, shorter distances may compromise the natural feel of a presentation. Another practical requirement for the operation environment is the need for two projection spaces or screens: one for the presenter's visual materials and the other for displaying the virtual audience. Additionally, to prevent strong echo reflections, it is recommended that the room not be completely empty. A normally equipped classroom provides sufficient echo dampening. In mostly empty rooms, the use of curtains or audio-absorbing foam is advised. An illustration of this environment is presented in Figure 8.

3.6 Licence

Though inspired by the functionality of prior systems (Schneider et al., 2015; Ochoa et al., 2018), OpenOPAF has been developed anew, using a distinct set of software dependencies. It does not reuse code from existing systems, so it does not infringe on their copyright. The code, 3D models, and related documentation are released under the General Public License (GPL) version 3[10]. This licence permits the use, modification, and distribution of the software, including for commercial purposes, under the condition that any alterations or redistributions of the software make the source code available and remain under the GPL licence.

[10]GPL v3: https://www.gnu.org/licenses/gpl-3.0.en.html

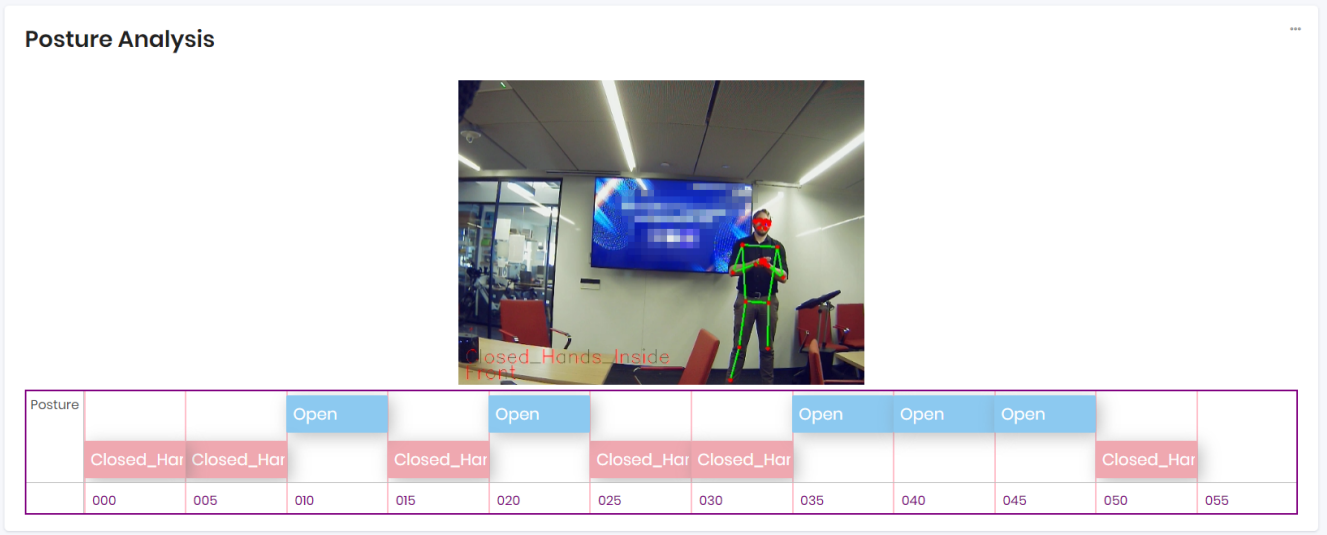


Figure 5. *Example of Posture feedback. Users can click on timeline segments to view specific parts of the video.*

4. Evaluation

To be a successful reverse-engineered version of existing OPAFs, OpenOPAF needs to perform similarly to the systems it is based on. To assess its performance, we conducted a small-scale evaluation focusing on three dimensions previously used to evaluate OPAF systems in the literature: feature extraction accuracy (Kurihara et al., 2007; Batrinca et al., 2013), learning gains as measured by the system (Damian et al., 2015; Ochoa & Dominguez, 2020), and presenters' perceptions (Schneider et al., 2015; Trinh et al., 2017). While the RAP system's learning gains, as measured by human experts, have also been evaluated in a controlled experiment (Ochoa, 2022a), this type of evaluation is beyond the scope of this work. Additionally, the RAP's controlled experiment primarily confirmed the results of the learning gains measured by the system.

These evaluations aim to provide insights into the following three questions:

1. Does OpenOPAF extract features of the captured modalities with accuracy comparable to previous OPAFs?
2. Do presenters using OpenOPAF achieve learning gains, as measured by the system, similar to those obtained with previous OPAFs?
3. Is the perception of presenters using OpenOPAF as positive as those found for other OPAFs?

It is important to note that the objective of this evaluation is not to provide a definitive and statistically significant answer to these questions but rather to offer initial evidence regarding OpenOPAF's performance. To obtain rigorous responses to these questions, OpenOPAF would need to be deployed in an ecologically valid setting and subjected to a randomized control trial. The following subsections present the methodology used in each dimension of the evaluation, the results obtained, and their comparison with those from previous OPAFs.

To conduct this evaluation, we recruited participants who were graduate students in education and related fields living in New York City, New York, USA. All participants were compensated 100 USD for their participation. After signing an IRB-approved informed consent agreement, participants were instructed to prepare a 5-minute presentation, on a topic that they had already presented elsewhere, to be performed twice, with sessions spaced at least one week apart. In both instances, participants were asked to (1) familiarize (or re-familiarize) themselves with the OpenOPAF system; (2) complete a paper-based questionnaire surveying their prior oral presentation experience, and in the second session, to recall feedback received in the first session; (3) deliver a 5-minute presentation using OpenOPAF; (4) review the feedback report generated by the system; and (5) participate in a brief interview to discuss their experience. We collected audio and video recordings of their presentations, images of their slides, questionnaire responses, interview notes, automated feedback reports, and features extracted by the system. All instruments used (questionnaires and interview questions) are available in this work's repository[11]. Given the nature of this evaluation, where rigorous statistical claims are not feasible, the instruments were tested only for comprehension with two individuals, similar to the recruited participants and external to the research team. The instruments were not statistically validated to measure any specific construct.

[11] Questionnaires and Interviews: https://osf.io/svc4p/?view_only=3492aab2784e45b893274851e9c714a1

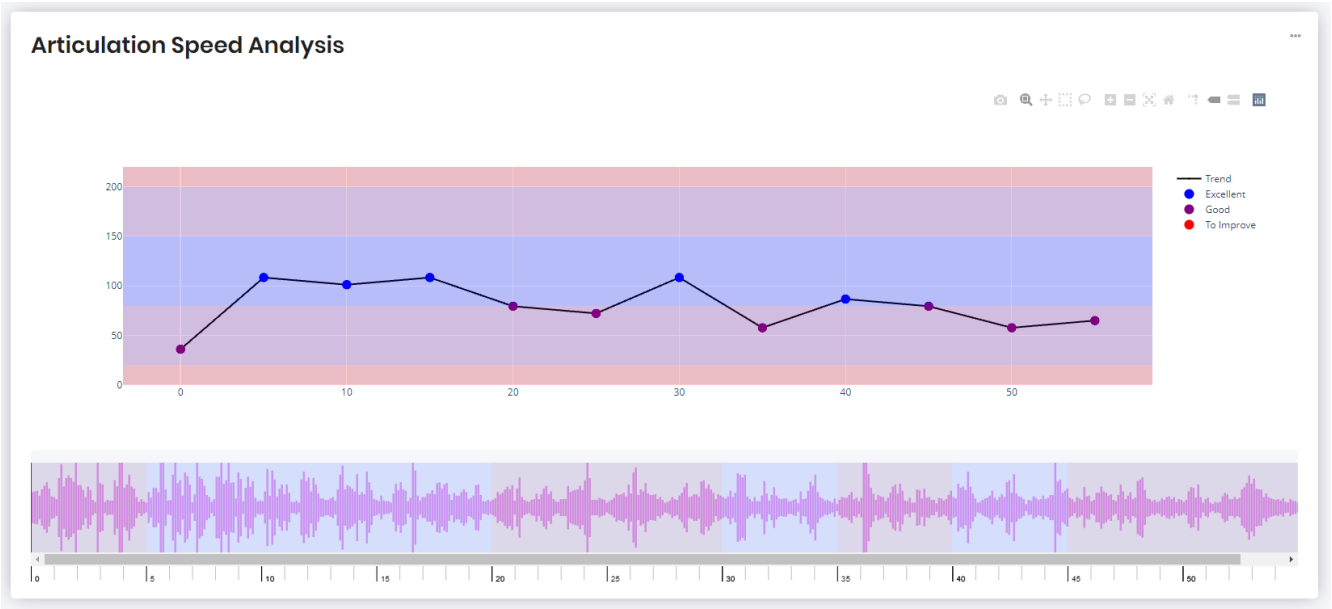


Figure 6. Example of Articulation Speed feedback. Users can click on the waveform to listen to speech segments.

A total of 12 participants completed two recording sessions, comprising 11 females and 1 male, aged between 25 and 43 years. There were four native English speakers and eight non-native speakers. This was a small convenience sample and is not claimed to be representative of the entire target population. The results obtained from this group of participants serve only as an indicator that OpenOPAF technically operates as intended, not as inferential proof of its effectiveness in real educational settings with a more diverse population.

Within the initial session's questionnaire, participants self-assessed their oral presentation skills, with seven rating themselves as "Advanced," two as "Intermediate," two as "Beginner," and one not responding. Regarding their prior experience, one reported "Extensive experience," five "A lot of experience," three "Some experience," and two "Little experience," and one did not respond. Additionally, an open-ended question about challenges faced in previous presentations revealed common themes of nervousness, speaking too quickly, and difficulties presenting in English (for non-native speakers). Given that nine out of 11 participants reported being at an advanced or intermediate level, and that nine out of 11 also cited at least some experience in presenting, it is reasonable to assume that most of them were not beginners.

4.1 Feature Extraction Accuracy

To answer the first evaluation question, we assessed the technical accuracy of the feature extraction by comparing the algorithm outputs with the responses provided to human coders. To evaluate the accuracy of the feature extraction algorithms, the audio and video recordings, as well as individual slides, were sampled. The sampling technique involved randomly selecting 20 instances in which the feature extractor had assigned a specific value. For example, 20 random frames classified as "Open" for body posture were chosen. If fewer than 20 instances existed for a given value, all available instances were selected. For instance, only five slides had an indeterminable font size, labelled as "None"; thus, all five slides were included in the sample. Refer to Table 3 for the final number of samples per modality.

The selected instances—video frames for video-based modalities, 5-second audio excerpts for audio-based modalities, and individual slides for slide-based modalities—were then coded by two independent human researchers. Both coders were graduate students in educational technologies, accustomed to presenting at seminars and conferences and attending other people's presentations. They were not involved in the authorship of this work. To code, they used the same categories available to the feature extractors. For example, the gaze direction extractor categorizes each frame with one of seven possible values: "Front," "Up," "Down," "Right," "Left," "Back," and "None." Human coders were instructed to assign one of these seven values to the sampled frames. The full code book used is available in an open repository[12]. Upon completion of coding, the kappa inter-rater reliability metric was calculated for each feature to assess agreement levels between coders. Instances with discrepancies were discussed to reach a consensus on the ground truth. Finally, the algorithm's assigned values were compared to this ground truth to calculate the percentage of agreement.

Table 3 presents the results for each feature. Four modalities (Gaze Direction, Voice Volume, Slides' Text Length, and Font

[12]Code book repository: https://osf.io/tqkjb/?view_only=ed791b94a62e4528a5123438e875dff5

ISSN 1929-7750 (online).

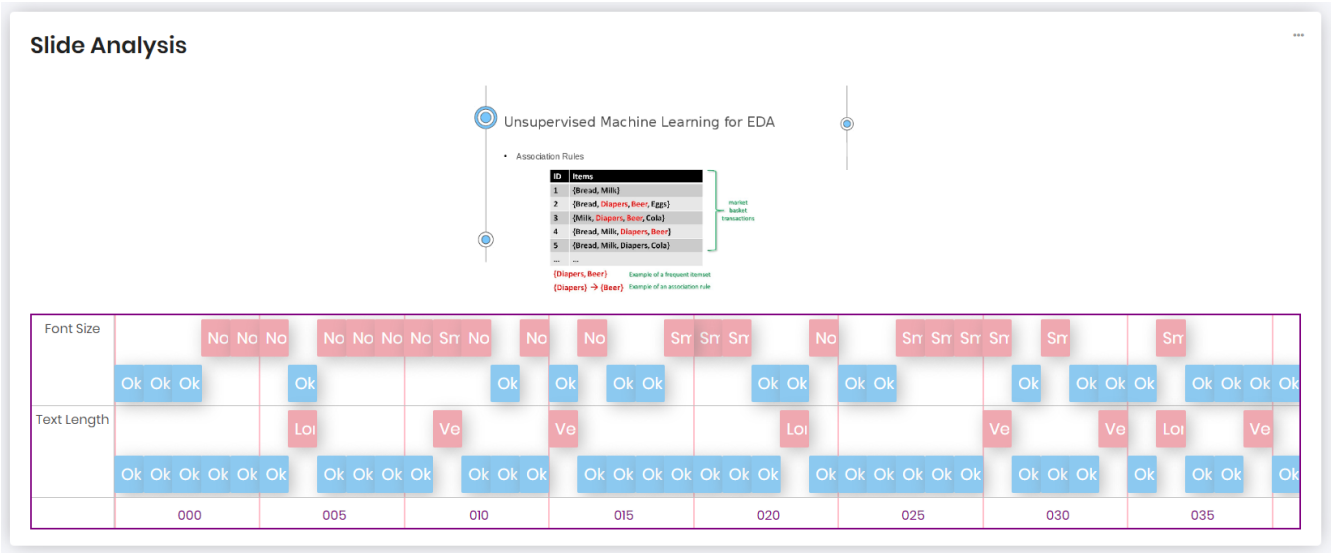


Figure 7. Example of Text Length and Font Size feedback. Clicking on extracted feature values displays the corresponding slide for reflection.

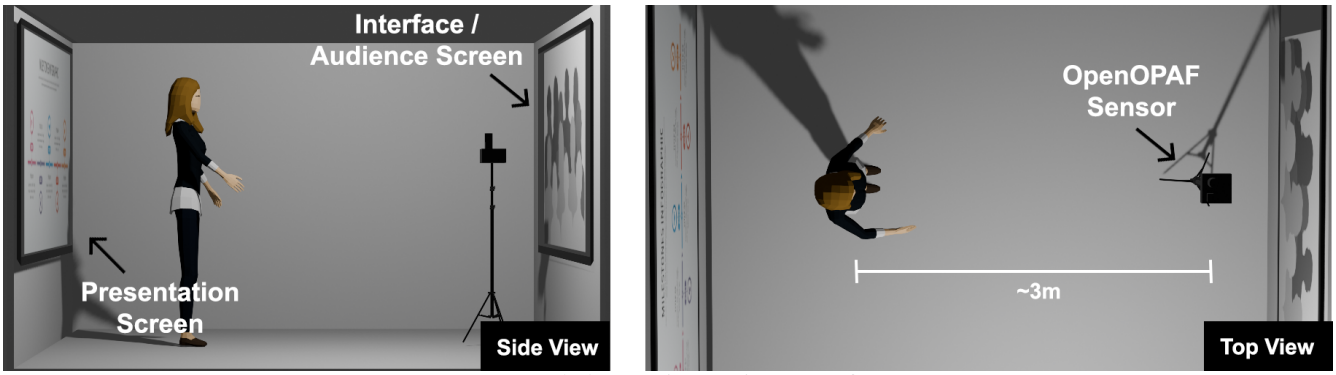


Figure 8. Recommended operation environment for OpenOPAF.

Size) achieved an agreement higher than 80%, while three modalities (Body Posture, Articulation Rate, and Filled Pauses) fell within the high 60s to low 70s range. Notably, the performance of the extractors does not correlate with the number of potential values but rather with the subjectivity involved in the task. This is evidenced by lower kappa values for the same modalities. For instance, a closer look at the Body Posture results reveals that the OpenOPAF system tends to classify a posture as non-open more frequently than human coders, who might consider it open. Additionally, many posture classification errors coincided with instances of coder disagreement, highlighting the challenge of managing subjectivity in the current rule-based system.

For comparison, Table 3 also includes technical accuracy results from the most recent OPAF systems, RAP (Ochoa et al., 2018) and RoboCOP (Trinh et al., 2017). The RAP system reports agreement between automatic extraction and human raters using three possible values for modality: “Bad,” “Medium,” and “Good.” The RoboCOP system reports F1 values across three categories for Filled Pauses and Articulation Rate, and two for Gaze Direction. Although F1 and Percentage of Agreement are calculated differently, both metrics similarly reflect system performance. Based on this initial comparative analysis, it seems that OpenOPAF’s feature extraction performance is on par with, if not better than, other OPAF systems.

4.2 Learning Gains

To address the second question, we analyzed the learning gains associated with the use of the system by measuring changes in various presentation skills detected by OpenOPAF across two consecutive recording sessions. OPAF systems are typically evaluated based on learning gains as measured by the system itself, even when the system’s scoring may not be perfectly accurate. This approach is often preferred over evaluations by human graders for two key reasons. First, obtaining an objective measure from human graders would require them to evaluate every individual frame or 5-second audio segment of the presentation video for each specific skill, resulting in an overwhelming amount of work, even for a small-scale study. Second, a more subjective but holistic assessment could involve human raters providing scores for each skill after viewing the entire

ISSN 1929-7750 (online).

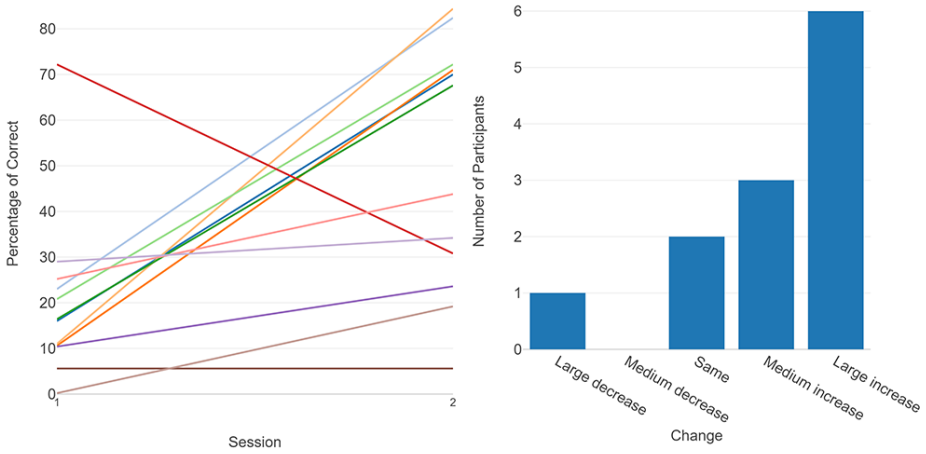
Table 3. Evaluation of the feature extraction accuracy.

Modality	Values	Samples	IRR—Kappa	Agreement	RAP Agreement	RoboCOP F1
Body Posture	8	150	0.76	69%	65%	–
Gaze Direction	7	135	0.91	83%	78%	84%
Voice Volume	2	40	0.89	93%	71%	–
Articulation Rate	3	65	0.69	74%	–	46%
Filled Pauses	3	32	0.87	72%	75/%	59%
Slides' Text Length	3	60	0.95	97%	46%	–
Slides' Font Size	3	65	0.93	88%	44%	–

presentation. However, human graders have shown inconsistency and unreliability in accurately assessing individual skills in oral presentations (Abdulkadir et al., 2021; Moothedath, 2024). As a result, human evaluations have only been used to assess OPAFs when the evaluation corresponds directly to the actual scores students receive in a course, which is valuable in its own right (Ochoa & Dominguez, 2020).

The scores for this evaluation were calculated based on the percentage of instances in which the system detected recommended or correct behaviours. For instance, if a participant's Gaze Direction was estimated as "Front" in 450 out of 500 recorded frames, their score for that modality would be 90%.

(a) Body Posture—Learning gains.



(b) Gaze Direction—Learning gains

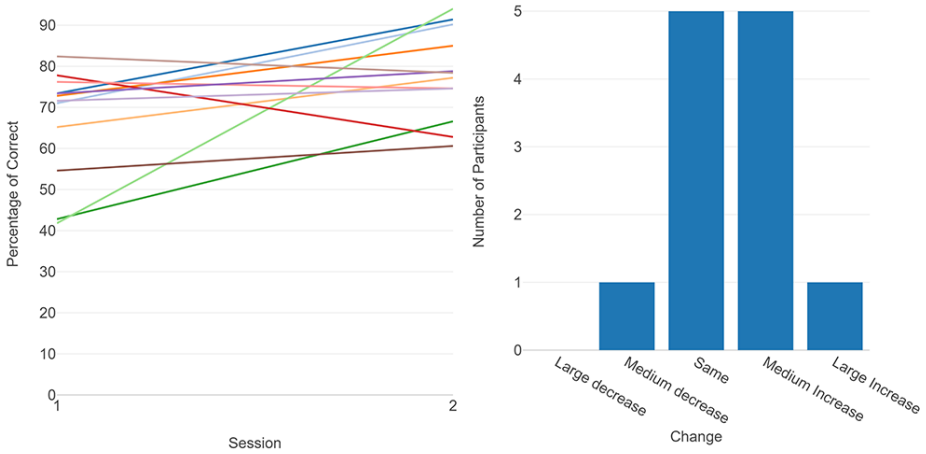
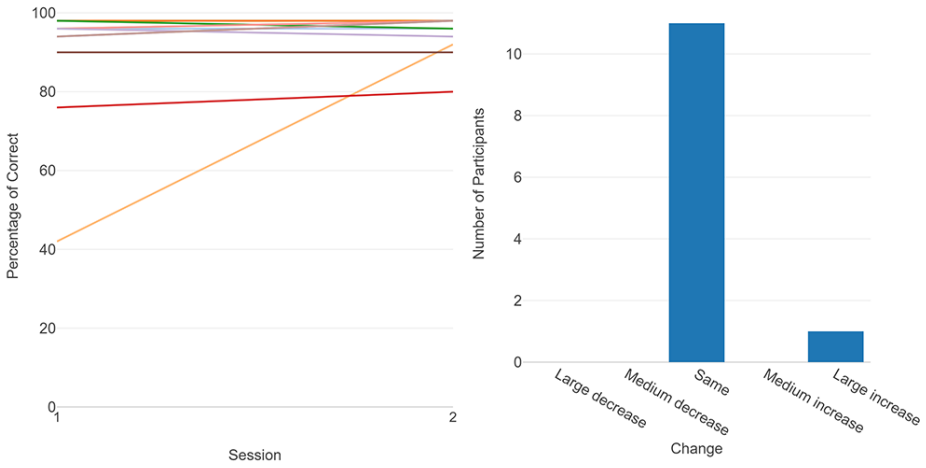
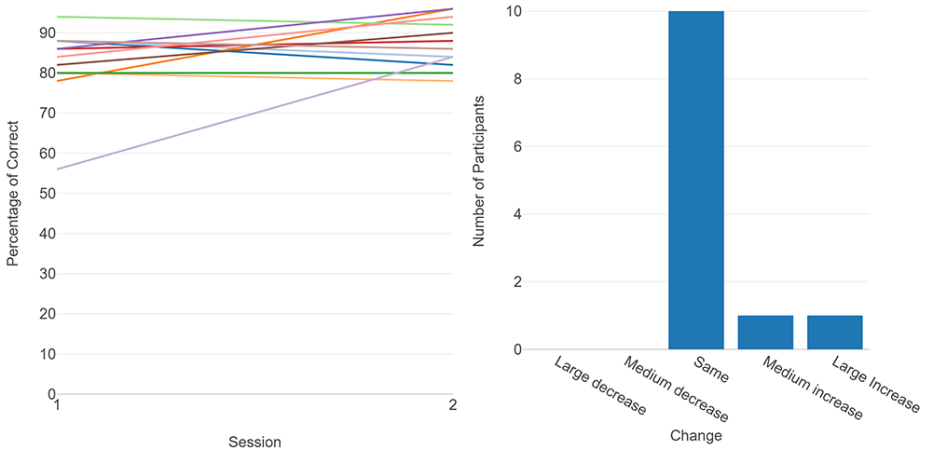


Figure 9. Learning gains for video-based modalities: (a) Body Posture, (b) Gaze Direction. In the parallel coordinates graph, each coloured line represents a different participant. The bar chart represents the aggregated change.

(a) Voice Volume—Learning gains.



(b) Articulation Rate—Learning gains



(c) Filled Pauses—Learning gains

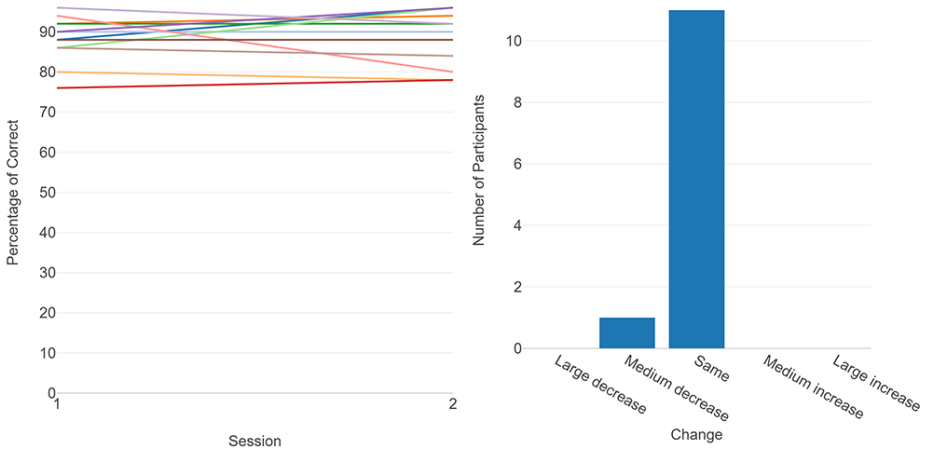


Figure 10. Learning gains for audio-based modalities: (a) Voice Volume, (b) Articulation Rate, and (c) Filled Pauses. In the parallel coordinates graph, each coloured line represents a different participant. The bar chart represents the aggregated change.

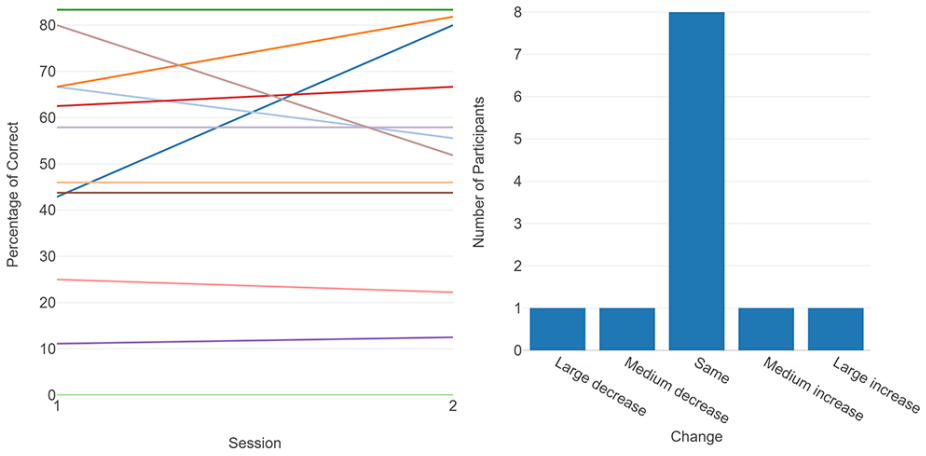
For Body Posture (see Figure 9a), a significant increase was noted for most participants from the first to the second

ISSN 1929-7750 (online).

presentation sessions. Initially, many participants scored low (30% or lower). In the second session, six participants showed a large increase (more than 25 percentage points), three had a moderate increase (between 10 and 25 percentage points), two remained about the same (within 10 percentage points), and one's score decreased notably. Statistical analysis revealed that scores in the second session ($M = 50.4, SD = 7.9$) were significantly higher ($t(11) = -3.16, p = .009$) than in the first session ($M = 20.0, SD = 5.3$). For Gaze Direction (see Figure 9b), all students initially received mid to high scores. Except for one, all participants either maintained (5) or increased (6 moderate, 1 large) their scores, with a less pronounced yet statistically significant positive change ($t(11) = -2.23, p = 0.047$), improving from a first session score ($M = 70.0, SD = 13.4$) to a second session score ($M = 77.9, SD = 11.0$).

The Voice Volume modality analysis (see Figure 10a) yielded inconclusive learning gains due to the high initial scores ($M = 89.5, SD = 16.3$). Most students scored similarly high in the second session ($M = 95.0, SD = 5.3$), though notably, the one student with a low initial score (42) improved significantly in the second session (91). The lack of statistical gains between sessions ($t(11) = -1.25, p = 0.24$) is attributed to generally good voice volume among participants. The Articulation Rate (see Figure 10b) and Filled Pauses (see Figure 10c) modalities followed a similar pattern, with initial high scores leading to high second-session scores and no statistically significant differences, indicating high initial vocal performance levels.

(a) Text Length—Learning gains



(b) Font Size—Learning gains

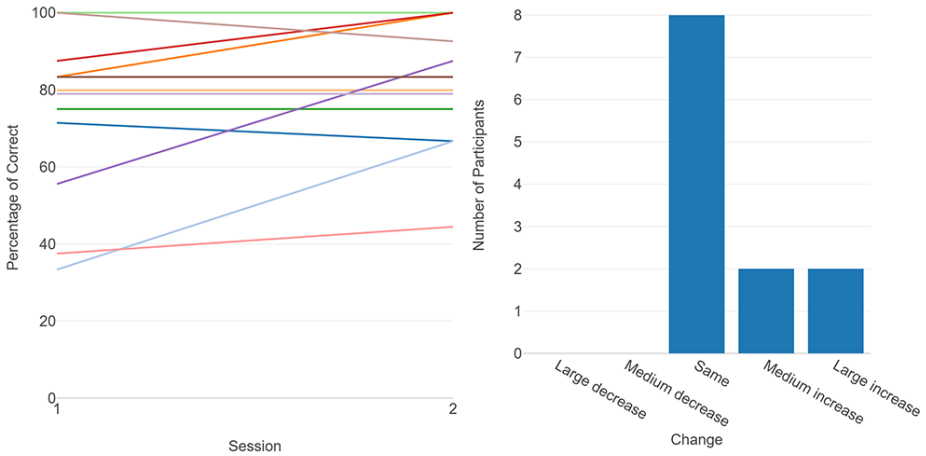


Figure 11. Learning gains for slide-based modalities: (a) Text Length, and (b) Font Size. In the parallel coordinates graph, each coloured line represents a different participant. The bar chart represents the aggregated change.

For slide-based modalities, scores varied more. Text Length (see Figure 11a) had a broad distribution in the first session ($M = 49.8, SD = 26.1$), with most students maintaining similar scores in the second session ($M = 50.1, SD = 27.2$), showing no statistical difference between sessions ($t(11) = -0.30, p = .77$). Font Size (see Figure 11b) also showed consistent average

scores between the first ($M = 73.8, SD = 21.5$) and second sessions ($M = 81.3, SD = 16.7$), without statistically significant differences ($t(11) = -1.89, p = .084$), due to minimal changes made by participants to their slides.

Overall, the analysis of learning gains from our study is consistent with findings from the most comprehensive controlled evaluation of OPAF systems to date (Ochoa & Dominguez, 2020). In their findings, they also identified significant improvements primarily in scenarios where the initial performance was universally low (such as with Gaze Direction) or when focusing specifically on lower-performing individuals (for Body Posture and Filled Pauses). These observations suggest that OpenOPAF similarly impacts the development of presentation skills among presenters as evaluated by the system itself.

4.3 Students' Perceptions

To gain insights into participants' perceptions and address the third evaluation question, we assessed the user experience through post-session interviews with presenters. All participants were surveyed and interviewed following their interactions with the OpenOPAF system. The answers to multiple-choice and open-ended questions in these surveys and interviews, combined with responses from the pre-presentation questionnaire in session 2, provided insights into the system's perceived usefulness and the impact of its feedback. The distribution of the answers can be seen in Figure 12. We analyzed five main aspects related to the report's accuracy, usefulness, and impact, and the participants' perceived personal improvement:

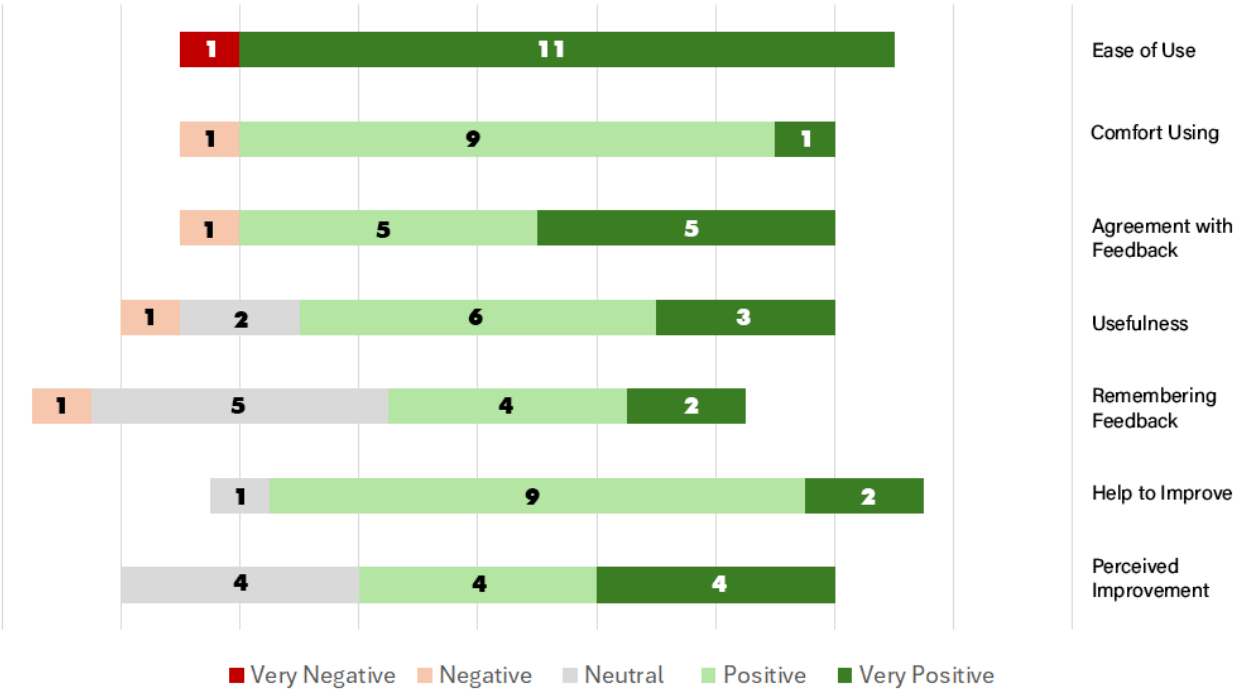


Figure 12. Responses of participants to multiple-choice questions in the questionnaires.

- **Ease of use:** In response to the Yes/No question, "Do you think our system is easy for you to use during the presentation?", all but one participant answered affirmatively. The dissenting participant cited difficulties maintaining eye contact with the audience while viewing slides on the laptop due to the room configuration in their response to the follow-up question, "If no, what problems did you encounter while using the system?". After the first session, in response to the 5-level Likert scale question, "How comfortable are you with the system?", two out of 12 participants rated their comfort using the system as "Extremely Comfortable," nine as "Somewhat Comfortable," and one as "Somewhat Uncomfortable." This improved in the second session, with eight participants responding "Extremely Comfortable" and four "Somewhat Comfortable." These results suggest a positive perception of the system's user-friendliness, which increases with familiarity (Davis, 1989).

- **Accuracy of the report:** When asked about their agreement with the feedback in the 5-level Likert scale question, "How much do you agree with the feedback on areas of improvement?", five participants chose "Strongly Agree," and another five chose "Somewhat Agree." One participant selected "Somewhat Disagree," and one did not respond. In a follow-up

ISSN 1929-7750 (online).

open question during the interview, "Do you agree or disagree with these suggestions in the report?", the disagreeing participant explained that their disagreement mainly pertained to the system's inability to detect nuanced presentation aspects, such as different types of pauses. Despite this, eight participants mentioned that they did not expect the high level of accuracy from the system when responding to the open-ended interview question, "What expectations did you have before using the system?". In the second session, in response to the open-ended question, "Did the report confirm those [improvement] changes?", all but one confirmed that the system captured their improvements. These results indicate that, despite the system's limitations, participants found it to accurately measure targeted modalities.

- **Usefulness of the report:** In response to the 5-level Likert scale question, "Will this report's feedback be useful for improving your presentation skills?", participants rated the report's usefulness for enhancing presentation skills as follows: three rated it "Extremely Useful," six "Very Useful," two "Moderately Useful," and one "Slightly Useful." No participant found the report "Not at all useful." In a follow-up open-ended interview question, "Based on the report you received, how do you feel about the feedback on areas of improvement?", participants mentioned that feedback on Body Posture (four participants) and Gaze Direction (five participants) was especially useful, aligning with learning gains analysis that showed improvements in these areas.

- **Impact of the report:** When asked about their recall of the first session's feedback with a 5-level Likert scale question, two participants said they remembered "All of it," four "A lot," five "A moderate amount," and one "A little." Before the second session, when asked with a 5-level Likert scale question, "Do you think the feedback from the first report helped you improve your presentation skills?", most participants (nine) responded "Probably yes," while two responded "Definitely yes," and one responded "Might or might not." Despite these responses and the compensation offered for their time, the effort to enhance presentations between sessions was minimal, with only half of the participants reporting that they changed their slides, which explains the non-significant learning gains for Text Length and Font Size.

- **Perceived improvement:** After the second session, responses to the 5-level Likert scale question question, "How do you feel your presentation went compared with the first time using the system?", were as follows: four responded "Much Better," four "Somewhat Better," and four "About the Same." In response to the open-ended question, "What did you improve the most?", participants mentioned "Posture" (10 participants), followed by "Gaze" (four), "Presentation Design" (three), and "Articulation Speed" (three). Two participants also noted increased "Confidence," a non-measured modality. Again, the perception aligns with the measurement, even when the improvements were not statistically significant.

These results confirm the positive perceptions of users toward systems like OpenOPAF, consistent with findings from previous perception evaluations (Schneider et al., 2015; Trinh et al., 2017; Ochoa et al., 2018; Ochoa & Dominguez, 2020). They underscore the demand for and potential of broader implementation of automated feedback systems for oral presentations in educational settings, with OpenOPAF marking a significant step toward expanding access to such systems.

4.4 Evaluation Conclusions

The evaluation of OpenOPAF provided valuable initial insights into the system's technical performance and user perceptions. The findings suggest that OpenOPAF performs comparably to other OPAF systems in terms of feature extraction accuracy and that users generally perceive the system positively, especially in terms of ease of use and the usefulness of its feedback.

However, it is important to note that the evaluation was conducted in a controlled, non-ecologically valid setting, where participants may have lacked intrinsic and extrinsic motivations to improve their presentation skills to the extent they might in a real educational environment. This limitation likely resulted in lower observed learning gains and less significant improvements in presentation skills than might be expected in a more authentic context.

In summary, while the current evaluation offers encouraging evidence of OpenOPAF's capabilities, the results should be interpreted as preliminary. Future studies conducted in more realistic educational contexts will be essential to fully understand the system's effectiveness and its potential to support students' presentation skills development in real-world scenarios.

5. Recommendations for Adoption and Improvement

As highlighted in the Introduction, the primary goal in sharing the design and components of the OpenOPAF system is threefold: to provide educational practitioners with a tool ready for facilitating the acquisition of oral presentation skills, to offer educational technologists an adaptable platform for various contexts, and to supply researchers with a unified platform for testing new automated multimodal feedback approaches. Here, we offer adoption recommendations tailored to these three use cases.

ISSN 1929-7750 (online).

5.1 Reuse As-Is

OpenOPAF includes all necessary elements for educational practitioners to implement it in a multi-week presentation practice and feedback cycle. Currently, the system best serves novice presenters still mastering skills such as maintaining open posture, engaging with the audience, speaking clearly at an acceptable pace, minimizing filler words, and designing readable slides. Evaluation results from this and previous studies (Ochoa & Dominguez, 2020) indicate that using OpenOPAF, and the current generation of OPAFs, provides no significant benefit (nor harm) to presenters who are already competent in these areas. For example, the voice volume and articulation rate of the current participants were already high enough to achieve top scores during the first and second sessions. As such, no valuable feedback was provided for these skills beyond confirming that they had mastered them.

Though designed for independent use by presenters, OpenOPAF can also enhance classroom activities. Instructors might explain and augment system feedback, offering personalized improvement suggestions. For classroom use, presentations could be shortened to 1 or 2 minutes to pinpoint common errors efficiently. We only recommend using the system for non-graded formative evaluation. The feature extraction algorithms could be easily fooled by individuals who want to game the system. Additionally, the thresholds used in the report are intentionally set to low values to boost presenter confidence and provide a positive experience, rather than to offer a strictly accurate performance assessment.

5.2 Adaptation

Educational technologists, particularly those capable of modifying OpenOPAF's hardware and software, can adapt the system to fit different learning contexts. Adaptations might include adjusting it for group presentations, altering report formats and thresholds, or integrating the hardware into a dedicated training space.

OpenOPAF was designed with adaptability in mind across several dimensions:

- **Target group expertise:** Adjusting the grading scheme allows OpenOPAF to cater to varying expertise levels, not just beginners. It's crucial to recognize the limits of such adjustments; for instance, setting appropriate increased gaze direction thresholds for intermediate presenters can work, but expecting 100% audience engagement is unrealistic, even for experts.
- **Report format and delivery:** The current report format is one of many possibilities. Changes could include emailing reports to presenters (and instructors, with consent), incorporating targeted recommendations, linking educational resources for improvement areas, and tracking progress over time.
- **Adding modalities:** OpenOPAF's modular nature simplifies the addition of new modalities, such as employing speech recognition and natural language processing to evaluate content relevance, as demonstrated in systems like RoboCOP (Trinh et al., 2017).
- **Additional functionality:** Access to the source code enables the addition of features like creating a Q&A session with the virtual audience by using speech-to-text conversion, in conjunction with large language models (LLMs), to generate audience questions post-presentation, and then text-to-speech synthesizers.

OpenOPAF's flexibility makes it a valuable resource for educational technologists, allowing for continual evolution to meet different educational needs and technological advancements. It is hoped that enhancements to OpenOPAF will be shared with the community, following the open-sharing ethos of its initial release.

5.3 Research

Researchers at the intersection of learning analytics and oral presentation skill development can leverage OpenOPAF to expedite testing of new ideas without the need to build a tool from scratch. Potential research directions include the following:

- **Feature extraction:** OpenOPAF's modular structure facilitates the evaluation and comparison of new feature extraction algorithms, enabling the direct assessment of their impact on presenters' learning gains.
- **Information delivery:** The system offers a unique opportunity to design and test different feedback interfaces and study how information uptake translates into behavioural changes, broadening our understanding of learning analytics' practical use.
- **Oral presentation skill acquisition:** With regular use over time, OpenOPAF provides detailed measurements for studying how novices develop presentation skills, potentially supporting or inspiring new theories on acquiring public speaking expertise.

OpenOPAF's open-source nature fosters a collaborative research environment, encouraging the sharing of discoveries and advancements with the wider community. This collaboration not only enhances the tool but also enriches the collective knowledge on effective presentation training.

6. Conclusions

This paper introduced the design, implementation, and evaluation of OpenOPAF, a system designed to provide automated feedback for oral presentations. Drawing inspiration from similar existing systems, our evaluation demonstrates that OpenOPAF performs comparably in both technical and pedagogical aspects. Although OpenOPAF does not introduce functional improvements over the current state of the art, its software and hardware specifications are openly shared with the learning analytics community for use, adaptation, and enhancement.

The open and flexible nature of OpenOPAF has the potential to make oral presentation feedback mechanisms more accessible. By offering a customizable platform tailored to various educational needs and goals, OpenOPAF promotes innovative teaching approaches and nurtures a culture of ongoing enhancement and collaboration. Future developments for OpenOPAF may include broadening its features, enhancing the precision of current modalities, and devising new feedback delivery methods to further engage and benefit learners. Additionally, exploring how the system can be incorporated into diverse learning environments presents an opportunity to make this type of system an integral part of the educational landscape.

Acknowledgements

The authors want to acknowledge the pioneering work of all the researchers and practitioners that designed, implemented, and evaluated previous iterations of OPAF systems that heavily guided and inspired OpenOPAF.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors declared no financial support for the research, authorship, and/or publication of this article

References

Abdulkadir, M. S., Rathnayaka, R., Kodithuwakkuge, V., & Beneragama, C. (2021). Reliability of assessing oral presentations by the university professionals. *International Journal of Research and Innovation in Social Science*, *5*(9), 378–383. https://doi.org/10.47772/IJRISS.2021.5912

Alley, M., & Robertshaw, H. (2004). Rethinking the design of presentation slides: Creating slides that are readily comprehended. In *Proceedings of the ASME International Mechanical Engineering Congress and Exposition* (ASME 2004), 13–19 November 2004, Anaheim, California, USA (pp. 445–450, Vol. 47233). ASME. https://doi.org/10.1115/IMECE2004-61889

Batrinca, L., Stratou, G., Shapiro, A., Morency, L.-P., & Scherer, S. (2013). Cicero—towards a multimodal virtual audience platform for public speaking training. In R. Aylett, B. Krenn, C. Pelachaud, & H. Shimodaira (Eds.), *Proceedings of the International Workshop on Intelligent Virtual Agents* (IVA 2013)*, Lecture notes in computer science* (pp. 116–128, Vol. 8108). Springer. https://doi.org/10.1007/978-3-642-40415-3_10

Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International*, *5*(9/10), 341–347. https://www.fon.hum.uva.nl/paul/papers/speakUnspeakPraat_glot2001.pdf

Bull, P., & Frederikson, L. (2019). Non-verbal communication. In A. M. Colman (Ed.), *Companion Encyclopedia of Psychology* (pp. 852–872). Routledge. https://doi.org/10.4324/9781315542072

Castañer, M., Camerino, O., Anguera, M. T., & Jonsson, G. K. (2013). Kinesics and proxemics communication of expert and novice PE teachers. *Quality & Quantity*, *47*, 1813–1829. https://doi.org/10.1007/s11135-011-9628-5

Chan, V. (2011). Teaching oral communication in undergraduate science: Are we doing enough and doing it right? *Journal of Learning Design*, *4*(3), 71–79. https://doi.org/10.5204/jld.v4i3.82

Chong, E., Wang, Y., Ruiz, N., & Rehg, J. M. (2020). Detecting attended visual targets in video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR 2020), 13–19 June 2020, Seattle, Washington, USA (pp. 5396–5406). IEEE. https://doi.org/10.1109/CVPR42600.2020.00544

Chunduru, V., Roy, M., Dasari, R. N. S., & Chittawadigi, R. G. (2021). Hand tracking in 3D space using MediaPipe and PnP method for intuitive control of virtual globe. In *Proceedings of the 2021 IEEE Ninth Region 10 Humanitarian Technology Conference* (R10-HTC 2021), 30 September 2021–02 October 2021, Bangalore, India (pp. 1–6). IEEE. https://doi.org/10.1109/R10-HTC53172.2021.9641587

Clegg, H. R., Carpenter, T. M., Freear, S., & Cowell, D. M. (2022). An open, modular, ultrasound digitial signal processing specification. In *Proceedings of the 2022 IEEE International Ultrasonics Symposium* (IUS 2022), 10–13 October 2022, Venice, Italy (pp. 1–4). IEEE. https://doi.org/10.1109/IUS54386.2022.9957486

Coleman, G. R., & Salter, W. T. (2023). More eyes on the prize: Open-source data, software and hardware for advancing plant science through collaboration. *AoB Plants*, *15*(2), 13. https://doi.org/10.1093/aobpla/plad010

Damian, I., Tan, C. S., Baur, T., Schöning, J., Luyten, K., & André, E. (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* (CHI 2015), 18–23 April 2015, Seoul, Republic of Korea (pp. 565–574). ACM. https://doi.org/10.1145/2702123.2702314

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

De Jong, N. H., Pacilly, J., & Heeren, W. (2021). PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically. *Assessment in Education: Principles, Policy & Practice*, *28*(4), 456–476. https://doi.org/10.1080/0969594X.2021.1951162

De Jong, N. H., & Wempe, T. (2009). PRAAT script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390. https://doi.org/10.3758/brm.41.2.385

Debalaxmi, D., Vishwakarma, D. K., & Ranga, V. (2024). Analyzing yoga pose recognition: A comparison of MediaPipe and YOLO keypoint detection with ensemble techniques. In *Proceedings of the Third International Conference on Applied Artificial Intelligence and Computing* (ICAAIC 2024), 5–7 June 2024, Salem, India (pp. 1011–1017). IEEE. https://doi.org/10.1109/ICAAIC60222.2024.10574984

Dermody, F., & Sutherland, A. (2015). A multimodal system for public speaking with real time feedback. In *Proceedings of the 2015 ACM International Conference on Multimodal Interaction* (ICMI 2015), 12–16 November 2015, Tokyo, Japan (pp. 369–370). ACM. https://doi.org/10.1145/2993148.2998536

Deshmukh, O., Espy-Wilson, C. Y., Salomon, A., & Singh, J. (2005). Use of temporal information: Detection of periodicity, aperiodicity, and pitch in speech. *IEEE Transactions on Speech and Audio Processing*, *13*(5), 776–786. https://doi.org/10.1109/TSA.2005.851910

Domínguez, F., Eras, L., Tomalá, J., & Collaguazo, A. (2023). Estimating the distribution of oral presentation skills in an educational institution: A novel methodology. In J. Jovanovic, I.-A. Chounta, J. Uhomoibhi, & B. McLaren (Eds.), *Proceedings of the 15th International Conference on Computer Supported Education* (CSEDU 2023), 21–23 April 2023, Prague, Czechia (pp. 39–46). ScitePress Digital Library. https://doi.org/10.5220/0011853900003470

Donnell, J. A., Aller, B. M., Alley, M. P., & Kedrowicz, A. A. (2011). Why industry says that engineering graduates have poor communication skills: What the literature says. In *Proceedings of the ASEE Annual Conference and Exposition* (ASEE 2011), 26–29 June 2011, Vancouver, British Columbia, Canada. ASEE. https://doi.org/10.18260/1-2--18809

Dowhower, S. L. (1991). Speaking of prosody: Fluency's unattended bedfellow. *Theory into Practice*, *30*(3), 165–175. https://doi.org/10.1080/00405849109543497

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363. https://doi.org/10.1037/0033-295X.100.3.363

Fernández-Nieto, G. M., Echeverria, V., Martinez-Maldonado, R., & Shum, S. B. (2024). YarnSense: Automated data storytelling for multimodal learning analytics. In *Proceedings of the 2024 Data Storytelling and Learning Analytics Workshop* (DS-LAK 2024), 18–22 March 2024, Kyoto, Japan (pp. 124–138). CEUR. https://ceur-ws.org/Vol-3667/DS-LAK24_paper_3.pdf

Gan, T., Wong, Y., Mandal, B., Chandrasekhar, V., & Kankanhalli, M. S. (2015). Multi-sensor self-quantification of presentations. In *Proceedings of the 23rd ACM International Conference on Multimedia* (MM 2015), 26–30 October 2015, Brisbane, Australia (pp. 601–610). ACM. https://doi.org/10.1145/2733373.2806252

Kilag, O. K. T., Quimada, G. M., Contado, M. B., Macapobre, H. E., Rabi, J. I. I. A., & Peras, C. C. (2023). The use of body language in public speaking. *Science and Education*, *4*(1), 393–406. https://openscience.uz/index.php/sciedu/article/view/4847

Kurihara, K., Goto, M., Ogata, J., Matsusaka, Y., & Igarashi, T. (2007). Presentation sensei: A presentation training system using speech and image processing. In *Proceedings of the Ninth International Conference on Multimodal Interfaces* (ICMI 2007), 12–15 November 2007, Nagoya, Aichi, Japan (pp. 358–365). https://doi.org/10.1145/1322192.1322256

Li, Z., Jensen, M. T., Nolte, A., & Spikol, D. (2024). Field report for platform mBox: Designing an open MMLA platform. In *Proceedings of the 14th International Conference on Learning Analytics and Knowledge* (LAK 2024), 18–22 March 2024, Kyoto, Japan (pp. 785–791). ACM. https://doi.org/10.1145/3636555.3636872

Lui, A. K.-F., Ng, S.-C., & Wong, W.-W. (2015). A novel mobile application for training oral presentation delivery skills. In J. Lam, K. Ng, S. Cheung, T. Wong, K. Li, & W. F. (Eds.), *International Conference on Technology in Education* (ICTE 2015)*, Communications in computer and information science* (pp. 79–89, Vol. 559). Springer. https://doi.org/10.1007/978-3-662-48978-9_8

Martinez-Maldonado, R., Echeverria, V., Prieto, L. P., Rodriguez-Triana, M. J., Spikol, D., Curukova, M., Mavrikis, M., Ochoa, X., & Worsley, M. (2018). Multimodal transcript of face-to-face group-work activity around interactive tabletops. In *Proceedings of the Second Multimodal Learning Analytics across (Physical and Digital) Spaces* (CrossMMLA 2018), 6 March 2018, Sydney, New South Wales, Australia. CEUR. http://ceur-ws.org/Vol-2163/paper4.pdf

McCarthy, C., Pradhan, N., Redpath, C., & Adler, A. (2016). Validation of the Empatica E4 wristband. In *Proceedings of the IEEE EMBS International Student Conference* (ISC 2016), 29–31 May 2016, Ottawa, Ontario, Canada (pp. 1–4). IEEE. https://doi.org/10.1109/EMBSISC.2016.7508621

McGaghie, W. C., Issenberg, S. B., Cohen, M. E. R., Barsuk, J. H., & Wayne, D. B. (2011). Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Academic Medicine: Journal of the Association of American Medical Colleges*, *86*(6), 706. https://doi.org/10.1097/acm.0b013e318217e119

Moothedath, M. (2024). Reliability of rubrics in the assessment of clinical oral presentation: A prospective controlled study. *Journal of Education and Health Promotion*, *13*(1), 182. https://doi.org/10.4103/jehp.jehp_1016_23

Motavalli, S. (1998). Review of reverse engineering approaches. *Computers & Industrial Engineering*, *35*(1), 25–28. https://doi.org/10.1016/S0360-8352(98)00011-4

Nguyen, A.-T., Chen, W., & Rauterberg, M. (2015). Intelligent presentation skills trainer analyses body movement. In I. Rojas, G. Joya, & A. Catala (Eds.), *Proceedings of the 13th International Work-Conference on Artificial Neural Networks* (IWANN 2015)*, Advances in computational intelligence* (pp. 320–332). Springer. https://doi.org/10.1007/978-3-319-19222-2_27

Ochoa, X. (2017). Multimodal learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gasevic (Eds.), *The Handbook of Learning Analytics* (pp. 129–141, Vol. 1). SoLAR. https://doi.org/10.18608/hla17.011

Ochoa, X. (2022a). Multimodal learning analytics—rationale, process, examples, and direction. In C. Lang, G. Siemens, A. Friend Wise, D. Gasevic, & A. Merceron (Eds.), *The Handbook of Learning Analytics* (2nd ed., pp. 54–65). SoLAR. https://doi.org/10.18608/hla22.006

Ochoa, X. (2022b). Multimodal systems for automated oral presentation feedback: A comparative analysis. In M. Giannakos, D. Spikol, D. Di Mitri, K. Sharma, X. Ochoa, & R. Hammad (Eds.), *The Multimodal Learning Analytics Handbook* (pp. 53–78). Springer. https://doi.org/10.1007/978-3-031-08076-0_3

Ochoa, X., & Dominguez, F. (2020). 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. https://doi.org/10.1111/bjet.12987

Ochoa, X., Domínguez, F., Guamán, B., Maya, R., Falcones, G., & Castells, J. (2018). The RAP system: Automatic feedback of oral presentation skills using multimodal analysis and low-cost sensors. In *Proceedings of the Eighth International Conference on Learning Analytics and Knowledge* (LAK 2018), 7–9 March 2018, Sydney, New South Wales, Australia (pp. 360–364). ACM. https://doi.org/10.1145/3170358.3170406

Olechowski, A., Eppinger, S. D., & Joglekar, N. (2015). Technology readiness levels at 40: A study of state-of-the-art use, challenges, and opportunities. In *Proceedings of the 2015 Portland International Conference on Management of Engineering and Technology* (PICMET 2015), 2–6 August 2015, Portland, Oregon, USA (pp. 2084–2094). IEEE. https://doi.org/10.1109/PICMET.2015.7273196

Rios, J. A., Ling, G., Pugh, R., Becker, D., & Bacall, A. (2020). Identifying critical 21st-century skills for workplace success: A content analysis of job advertisements. *Educational Researcher*, *49*(2), 80–89. https://doi.org/10.3102/0013189X19890600

Schneider, J., Börner, D., Van Rosmalen, P., & Specht, M. (2015). Presentation Trainer, your public speaking multimodal coach. In *Proceedings of the 2015 ACM International Conference on Multimodal Interaction* (ICMI 2015), 9–13 November 2015, Seattle, Washington, USA (pp. 539–546). ACM. https://doi.org/10.1145/2818346.2830603

Schneider, J., Börner, D., Van Rosmalen, P., & Specht, M. (2016). Enhancing public speaking skills—An evaluation of the Presentation Trainer in the wild. In K. Verbert, M. Sharples, & T. Klobučar (Eds.), *Proceedings of the 2016 European Conference on Technology Enhanced Learning* (EC-TEL 2016)*, Lecture notes in computer science* (pp. 263–276). Springer. https://doi.org/10.1007/978-3-319-45153-4_20

Schneider, J., Romano, G., & Drachsler, H. (2019). Beyond reality—extending a presentation trainer with an immersive VR module. *Sensors*, *19*(16), 3457. https://doi.org/10.3390/s19163457

Subapriya, K. (2009). The importance of non-verbal cues. *Journal of Soft Skills*, *3*(2), 37–42. https://www.iupindia.in/609/IJSS_Non-Verbal%20Cues_37.html

Tanveer, M. I., Lin, E., & Hoque, M. (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* (IUI 2015), 29 March–1 April 2015), Atlanta, Georgia, USA (pp. 286–295). ACM. https://doi.org/10.1145/2678025.2701386

Thurneck, L. (2011). Incorporating student presentations in the college classroom. *Inquiry*, *16*(1), 17–30. https://files.eric.ed.gov/fulltext/EJ952023.pdf

Trinh, H., Asadi, R., Edge, D., & Bickmore, T. (2017). RoboCOP: A robotic coach for oral presentations. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(2), 1–24. https://doi.org/10.1145/3090092

Van Ginkel, S. (2019). *Fostering oral presentation competence in higher education* [Doctoral dissertation, Wageningen University]. https://doi.org/10.18174/476541

Van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2015). Towards a set of design principles for developing oral presentation competence: A synthesis of research in higher education. *Educational Research Review*, *14*, 62–80. https://doi.org/10.1016/j.edurev.2015.02.002

Van Ginkel, S., Laurentzen, R., Mulder, M., Mononen, A., Kyttä, J., & Kortelainen, M. J. (2017). Assessing oral presentation performance: Designing a rubric and testing its validity with an expert group. *Journal of Applied Research in Higher Education*, *9*(3), 474–486. https://doi.org/10.1108/JARHE-02-2016-0012

Williams van Rooij, S. (2011). Higher education sub-cultures and open source adoption. *Computers & Education*, *57*(1), 1171–1183. https://doi.org/10.1016/j.compedu.2011.01.006

Worsley, M., & Martinez-Maldonado, R. (2018). Multimodal learning analytics' past, present, and potential futures. In *Proceedings of the Second Multimodal Learning Analytics across (Physical and Digital) Spaces* (CrossMMLA 2018), 6 March 2018, Sydney, New South Wales, Australia (pp. 1–16, Vol. 2). CEUR. https://ceur-ws.org/Vol-2163/paper5.pdf

Yan, L., Zhao, L., Gasevic, D., & Martinez-Maldonado, R. (2022). Scalability, sustainability, and ethicality of multimodal learning analytics. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 13–23). ACM. https://doi.org/10.1145/3506860.3506862