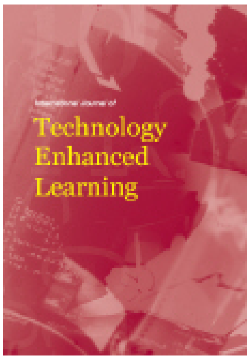


International Journal of Technology Enhanced Learning

ISSN online: 1753-5263 - ISSN print: 1753-5255
https://www.inderscience.com/ijtel

Enhancing presentation skills: key technical features of automated feedback systems - a systematic feature analysis

Stefan Hummel, Jan Schneider, Nina Mouhammad, Roland Klemke, Daniele Di Mitri

DOI: 10.1504/IJTEL.2025.10073166

Article History:

Received:	28 April 2025
Last revised:	05 May 2025
Accepted:	21 May 2025
Published online:	14 September 2025

Enhancing presentation skills: key technical features of automated feedback systems – a systematic feature analysis

Stefan Hummel*, Jan Schneider and Nina Mouhammad

DIPF | Leibniz Institute for Research and Information in Education,
Rostocker Str. 6, 60323 Frankfurt am Main, Germany
and
Goethe University Frankfurt,
Theodor-W.-Adorno-Platz 1,
60323 Frankfurt am Main, Germany
Email: s.hummel@dipf.de
Email: j.schneider@dipf.de
Email: n.mouhammad@dipf.de
*Corresponding author

Roland Klemke

TH Köln University of Applied Sciences,
Cologne Game Lab,
Schanzenstr. 28, 51063 Köln, Germany
Email: rk@colognegamelab.de

Daniele Di Mitri

DIPF | Leibniz Institute for Research and Information in Education,
Rostocker Str. 6, 60323 Frankfurt am Main, Germany
Email: d.dimitri@dipf.de

Abstract: Oral presentation automated feedback systems (OPAFs) aim to enhance public speaking by offering automated feedback. Despite their potential, most remain at the prototype stage and are not widely adopted in education. This study systematically analyses the implementation of functional features in existing OPAFs and investigates the evaluation methodologies used in published research. A scoring sheet, developed from expert input and supporting literature, was used to analyse 14 systems across 83 features and 12 additional aspects. Results show a low implementation rate of just 16%, with notable gaps in verbal-nonverbal congruency, adaptive feedback, and content structuring. Evaluation methods often focus on usability and user experience, while learning outcomes and pedagogical value are rarely addressed. Most studies rely on lab-based evaluations, limiting generalisability. The findings underline the need for improved feature integration, real-world testing, and collaboration with educators. Addressing these issues could support the transition from experimental tools to effective educational technologies.

Keywords: oral presentation automated feedback systems; OPAFs; public speaking; feature analysis; automated feedback; technology acceptance; system adoption.

Reference to this paper should be made as follows: Hummel, S., Schneider, J., Mouhammad, N., Klemke, R. and Di Mitri, D. (2025) 'Enhancing presentation skills: key technical features of automated feedback systems – a systematic feature analysis', *Int. J. Technology Enhanced Learning*, Vol. 17, No. 6, pp.1–25.

Biographical notes: Stefan Hummel began working as a Doctoral researcher in the Educational Technologies group at DIPF in 2022. His PhD research focuses on improving user acceptance of automated feedback systems for oral presentations. Prior to this, he studied Business Informatics at Technische Hochschule Mittelhessen, where he completed his Master's degree with a thesis on the application of data science methods to cryptocurrency markets. He also worked as an IT support member and was active in the academic self-government.

Jan Schneider is a senior researcher and developer for EduCS at DIPF. In 2017, he received his Doctorate from the Open University of the Netherlands on 'Sensor-based learning support'. He has published over 90 peer-reviewed scientific publications in academic journals and conference proceedings. His research work has received multiple awards at international conferences (EC-TEL 2014, EC-TEL 2015, ICMI 2015, LAK 2019). He completed the MILKI-PSY project as project coordinator at the DIPF. He worked as the scientific coordinator in the recently completed IWWB-PLUS project. He has also been Demo and Poster Chair at the EC-TEL conference in 2019, and in 2021 he was Workshop Chair.

Nina Mouhammad is a Doctoral researcher at the Leibniz Institute for Research and Information in Education (DIPF) in Frankfurt am Main, Germany. She holds a Master's in Computer Science with a minor in Human Sciences, primarily psychology, from TU Darmstadt. Integrating technology with education to enhance learning experiences is a key focus of her work. She aims to bridge the gap between technical innovation and practical application in educational settings. In her PhD research, she focuses on how software could support students with learning to create more effective content for their presentations.

Roland Klemke is a Professor of Game Informatics at the CGL and Associate Professor at the Faculty of Educational Science at the Open University of the Netherlands (Head of the Multimodal Learning Experiences Group). His research focuses on artificial intelligence for games, game engine architectures, development methods, serious games, game-based learning, augmented and mixed reality, and multi-sensor architectures. He is responsible for national and international research projects. He is a member of the German Informatics Society (GI), a Fellow of the Interuniversity Center for Educational Sciences (ICO), and the Research School Information and Knowledge Systems (SIKS).

Daniele Di Mitri leads the 'augmented feedback' group at DIPF and lectures at Goethe University Frankfurt. He earned his PhD from the Open University of the Netherlands in 2020. His research focuses on multimodal data, AI, and feedback in education. He is a Johanna Quandt Young Academy Fellow and winner of the 2018 Martin Wolpers Award. Daniele serves on the editorial board of Frontiers in Artificial Intelligence, Co-chairs the LAKathon series, and is active in the CrossMMLA special interest group of the Society for Learning Analytics Research.

1 Introduction

Effective communication is a cornerstone of success in both personal and professional life. Among the many facets of communication, public speaking stands out as an essential skill, particularly in the 21st century, where it plays a pivotal role in various scenarios such as job interviews (Patil et al., 2024), negotiations (Svendsen, 2022), and collaborative projects (Chan, 2011).

In the scientific community, the ability to convey research findings through compelling presentations can significantly impact a researcher's career and contribute to the advancement of the broader scientific enterprise. As Dave Rubenson aptly notes, a well-structured presentation is not just a 'data dump', but a coherent narrative that can help researchers organise their data, promote critical thinking, and gain valuable insights from their audience (Rubenson, 2021).

Acquiring public speaking skills, however, is not a straightforward process. It requires deliberate practice and training in various communication techniques, such as effective body language, tone modulation, and speech clarity (Kerby and Romine, 2009). Several methods are available for honing these skills, ranging from practicing alone in front of a mirror, which is a common but limited approach regarding feedback (Levasseur et al., 2004), to using video recordings, a widely used method that allows speakers to review and refine their presentations based on observed strengths and weaknesses (Zimmerman and Schunk, 1989). A more structured option are group courses. Their value is shown through approximately 1.3 million students in the USA alone, who enroll in basic communication courses each year (Morreale et al., 2016). Another effective practice, however, involves individualised instruction with direct feedback from teachers, allowing for personalised guidance and improvement (Heinicke et al., 2022). In recent years, technological advancements have introduced oral presentation automated feedback systems (OPAFs), as a modern tool for enhancing public speaking skills (Monteiro et al., 2024).

OPAFs are software-based systems that leverage various sensors, such as cameras and microphones, to provide automated feedback on a presenter's performance, with the goal of improving their communication skills (Edwards et al., 2018). Despite their potential of independently improving communication skills, OPAFs have not yet gained widespread adoption in educational settings and are often found only as research prototypes rather than fully integrated tools (Ochoa, 2022). This leads us to our main research question: Why have not OPAFs been widely implemented in educational contexts despite their promise?

To address this question, first, we need to identify relevant features that increase the usefulness of OPAFs. To find these features, we conducted a preliminary study where we interviewed experts to get important key aspects that are relevant for the OPAFs to cover. We then looked deeply at these key aspects in the literature, which allowed us to identify 83 functional features and 12 other relevant aspects.

This list of functional features led us to our first research question:

RQ1 To what extent have relevant functional features that affect the development of public speaking skills been implemented in existing OPAFs?

Moreover, to push the state-of-the-art of OPAFs in terms of research practices, we need to identify first how OPAFs have been evaluated, therefore our second research question is:

RQ2 How and on which criteria have these systems been evaluated, and what were the results from these evaluations?

To address these questions, we conducted a systematic feature analysis to identify the strengths and gaps in OPAF systems. First, we searched scientific databases using predefined search strings to find relevant OPAF systems published in scientific journals. Next, we developed a scoring sheet incorporating 83 functional features and 12 additional relevant aspects. We then provided these scoring sheets, along with the corresponding scientific papers, to 15 researchers, who evaluated the systems using the DESMET method. Finally, the evaluation results highlighted both the strengths and gaps of the analysed systems.

2 Background

In formal educational settings, a presentation consists of 4 parts: content research, message composition, slide design, and rehearsal. OPAFs have traditionally focused on the rehearsal aspect of a presentation. The rehearsal aspect covers categories like verbal and non-verbal communication behaviours (delivery), spoken content and structure (organisation and central message), feedback regarding supporting material, the way to provide feedback, providing best practices or examples (guidance), and expert opinion regarding the automated feedback (teacher involvement). Besides these important categories, OPAFs in general, try to improve presentation skills, which can be measured through studies that have a focus on learning gains. But also social effects when interacting with these types of systems, play a significant role. Rehearsing in general, is a concept of keeping the user in the loop, and therefore factors like usability, user experience, and technology acceptance are mandatory to keep up motivation. Besides all these mentioned factors, OPAFs are in a field where automated systems compete against human expertise, and therefore studies about the accuracy and performance of these systems are essential to figure out the strengths and weaknesses of these types of systems.

2.1 Delivery

The delivery of a presentation is composed of nonverbal-, verbal communication, and their consistency as well as time management. Nonverbal communication includes elements such as postures, gestures, and facial expressions, which significantly impact the quality of a presentation (Schneider et al., 2017). Schneider et al. (2017) identified several nonverbal features that affect presentation quality and grouped them into key categories. Verbal communication, on the other hand, involves elements such as vocal variety, articulation, and the effective use of language, all of which contribute to how well the message is conveyed to the audience. Schreiber et al. (2012) identified these verbal elements as critical components of effective public speaking and highlighted their importance in the overall delivery process. Gillis and Nilsen (2017) emphasised that consistency between verbal and nonverbal communication is a key factor in determining a speaker's credibility, particularly in contexts where the audience is evaluating the trustworthiness of the information being presented. Time management is also crucial, ensuring that the presentation fits within the allocated time (Schreiber et al., 2012). We examined the extent to which OPAFs accounted for these aspects in practice. For our feature analysis of OPAFs, we focused on the most common features reported in these studies for both nonverbal and verbal communication, as well as their consistency and also time management.

2.2 Organisation and central message

The organisation and central message of a presentation include the structure (introduction, middle, conclusion). As Schreiber et al. (2012) mentioned, for formal or academic presentations, the introduction should capture the audience’s attention and clearly introduce the topic and problem. The middle should present all main points logically and support them with evidence, while the conclusion should summarise the content and refer back to the main thesis. For our feature analysis of OPAFs, we examined whether the aspects listed were taken into account.

2.3 Supporting material

Supporting materials, such as visual aids and other presentation tools, play a crucial role in enhancing the clarity and impact of a presentation. As Schreiber et al. (2012) highlight, they not only support the delivery of the central message but also help sustain audience engagement. Similarly, Collins (2004) emphasises that the effectiveness of a presentation largely depends on how skillfully visual aids are employed to communicate with the audience. In our analysis of OPAF systems, we examined these aspects as highlighted by the authors.

2.4 Feedback

Feedback is critical in refining presentation skills and includes several subcategories: timing of feedback, corrective feedback, metacognitive/self-awareness feedback, and motivational feedback. Immediate feedback can be more effective in some contexts, while delayed feedback might be better suited in others, depending on the learning objectives (Mory, 2003). The nature of feedback also impacts its effectiveness; for instance, formative (corrective) feedback is essential just before the next practice opportunity (Keller, 1987). Motivational feedback, which encourages greater effort, engagement, and self-regulation, is also vital, as it can significantly enhance a speaker's confidence and willingness to improve (Hattie and Timperley, 2007). We evaluated whether OPAF systems effectively incorporated these feedback types, as described by the authors.

2.5 Guidance

Guidance is essential in the context of OPAF systems, particularly in onboarding users and helping them understand how to effectively use the system's features. Effective onboarding processes, like those discussed in Santos et al. (2024), are crucial for ensuring that users can adapt to and make the most of new environments and tools. In our analysis of OPAF systems, we examined whether the systems provided clear and intuitive guidance to users, facilitating their ability to receive and implement feedback effectively.

2.6 Teacher involvement

Teacher Involvement is a crucial aspect of OPAF systems. Concerns from experts were raised about the reliability of computer-generated feedback. For example, certain nonverbal cues might be misinterpreted by automated systems as negative when in some contexts, this could be perfectly acceptable. Therefore, it is important for OPAF systems to allow teachers to remain involved in the evaluation process to ensure that the feedback is holistic and contextual (Rodríguez-Triana et al., 2018).

Table 1 Criteria and their definition regarding the evaluation process of OPAFs

Criteria	*Description and measurement method*
Usability	‘Usability is the extent to which a system can be used by specific users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specified context of use’ (ISO, 2018). It can be measured, e.g., with a standardised questionnaire (Brooke, 1996) but also with techniques like the think-aloud protocol
Technology acceptance	Technology acceptance is the process by which individuals decide to adopt and use a technology based on perceived usefulness and perceived ease of use. It can be measured, e.g., with a standardised questionnaire (Venkatesh and Davis, 2000) or newer variations of this standardised questionnaire (Venkatesh and Bala, 2008)
User experience	‘User's perceptions and responses that result from the use and/or anticipated use of a system’ (ISO, 2019). It can be measured, e.g., with a standardised questionnaire (Laugwitz et al., 2008)
Accuracy and performance	Accuracy and performance in the context of automated feedback systems refer to the degree to which the system's output aligns with expert human judgment and predefined evaluation criteria (Hülsmann et al., 2017). Accuracy is typically assessed using inter-rater reliability metrics (e.g., Cohen’s Kappa, Krippendorff’s Alpha) to measure agreement between the system and human evaluators. Performance is evaluated through statistical measures such as precision, recall, and F1-score for classification tasks, as well as threshold-based analyses that track deviations beyond predefined limits (e.g., detecting excessive speech pauses, vocal intensity variations, or nonverbal cue discrepancies)
Learning gains	In educational research, learning gain is often defined as the ‘distance travelled’ or the difference between the skills, competencies, content knowledge, and personal development demonstrated by students at two points in time (McGrath et al., 2021). These gains are often assessed by comparing pre – and post-intervention performance metrics
Social effects	Social effects refer to the ways in which a system simulates or facilitates social presence, social interaction, and audience-related dynamics, thereby influencing a user's motivation, engagement, self-perception, or anxiety in communicative contexts. This includes elements such as simulated audience feedback (e.g., virtual humans reacting), the perceived social presence of the system, its impact on performance anxiety or confidence, and the realism and immersion of socially relevant scenarios. While Biocca et al. (2003) critique existing definitions of social presence as too vague for empirical use, they provisionally describe it as the ‘sense of being with another’ in a mediated or virtual environment, or more broadly as the ‘sense of being together’. To assess the social effects in such systems, self-report questionnaires, such as those measuring public speaking anxiety or perceived presence, are commonly used in experimental settings

2.7 Evaluation of OPAFs

Evaluating OPAFs is essential to understand how effectively these systems achieve their intended goals. This requires assessing their technical performance as well as their pedagogical effectiveness. A comprehensive evaluation helps to identify gaps and areas for improvement to ensure successful integration into educational settings. Based on insights from the experts we interviewed, we have identified six key criteria that significantly impact the success of these systems (see Table 1).

3 Method

This study employed a multi-phase methodological approach to evaluate key features of OPAF systems. First, a scoring sheet was developed based on expert input from 13 specialists in presentation and communication skills, supported by relevant literature (see background section). This scoring sheet, which covers 83 functional features and 12 other relevant aspects, was then used to assess existing systems. To identify relevant OPAF systems, a systematic search was conducted across five major academic databases (see Table 2), leading to the selection of 14 systems based on predefined inclusion and exclusion criteria. Finally, a systematic evaluation was performed using the DESMET methodology in screening mode (Kitchenham et al., 1997), where 15 researchers independently assessed the selected systems. This structured approach ensured a comprehensive analysis of current OPAF capabilities, highlighting feature gaps and areas for improvement.

3.1 Scoring sheet creation and distribution

The scoring sheet was developed based on insights gathered from semi-structured interviews with 13 experts who specialise in lecturing on enhancing presentation and communication skills, alongside relevant literature (see background section) that supported the claims made by these experts. The experts from Germany, the USA, and Croatia, identified key factors influencing the effectiveness of oral presentations, including delivery, organisation, the use of supporting materials, feedback, guidance, and teacher involvement. These elements, along with other important aspects, such as system evaluation, formed the foundation of the scoring sheet.

The final scoring sheet contained 83 functional features and 12 other relevant aspects. Each feature was rated on a scale of:

- 0 not mentioned or implemented
- 1 partially mentioned or implemented
- 2 Fully implemented.

Additionally, a comment section allowed reviewers to provide detailed explanations for their ratings. The scoring sheet was distributed to 15 researchers who independently reviewed various OPAF systems based on these established criteria, ensuring a thorough and reliable evaluation process.

3.2 Systematic feature analysis of existing OPAF systems

Our methodological choice to conduct a systematic search procedure based on scientific papers was driven by the proprietary and closed nature of most OPAF systems, which inhibits external scientific testing, as Ochoa and Zhao (2024) pointed out.

A number of literature reviews (Sharma and Giannakos, 2020; Blikstein and Worsley, 2016; Worsley, 2018; Di Mitri et al., 2018) have been conducted in the broader field of multimodal learning analytics. Although the literature provides an overview of OPAFs (Ochoa, 2022), there is no systematic feature analysis specifically on OPAFs, especially with our selected criteria (see Table 2). Based on these facts, we applied our own search with the following inclusion and exclusion criteria.

- Inclusion criteria:
  - I1 Automated feedback systems to improve public speaking skills: the system must be an automated feedback system designed to support rehearsal-based practice for improving presentation or public speaking skills.
  - I2 Completeness of internal processing steps: the system must implement all core processing steps – from capturing sensor data (e.g., audio, video, motion) to analysing and generating automated feedback – without relying solely on human evaluation.
- Exclusion criteria:
  - E1 Systems without a clear pedagogical goal: systems that do not explicitly aim to enhance presentation or public speaking skills through rehearsal-based feedback will be excluded. This includes general communication analytics or visualisation tools.
  - E2 Conceptual or non-implemented systems (‘paper systems’): systems that exist only as theoretical frameworks or proposed models without a working prototype or empirical validation are excluded. This ensures that only systems with actual implementations are considered.
  - E3 Systems not mentioned in peer-reviewed scientific literature: systems that are purely commercial products without transparent, documented evaluations in academic publications have been excluded. This prevents proprietary tools from being included whose technical details and effectiveness cannot be verified.

After the first screening based on inclusion and exclusion criteria, the number of relevant literature dropped from 1,587 to 23 (see Table 3). However, based on the references to similar systems in these research papers, the number of relevant research papers increased from 23 to 45. It also turned out that many publications have different research goals, but for their research, they used the same ‘basic OPAF system’ (sometimes also with modifications to that basic system depending on the underlying research goal).

To offer a greater selection and avoid prioritising authors who published many research papers on the same basic OPAF system with small modifications compared to those who only published one or two research papers, we also made the decision to exclude similar variations of those OPAF systems.

Table 2 Search criteria regarding literature for OPAFs

Parameter	*Decision*	*Reason*
Databases	ACM digital library, IEEE Xplore, MDPI, Science Direct and Springer Link	These databases were also used by other researchers for a comprehensive literature review in this field (Ochoa, 2022)
Search string	(‘public speaking’ OR ‘presentation training’ OR ‘presentation feedback’) AND (‘coach’ OR ‘tutor’ OR ‘AI tutor’ OR ‘trainer’ OR ‘application’ OR ‘system’)	We aim to find publications that describe systems for improving presentation or communication skills. Derivatives where several factors are tackled should also be included
Search location	Title, abstract, and keywords	We believe that the keywords in the search string should be included in the title, abstract or mentioned as keywords
Time restrictions	no time restrictions (until November 2023)	Our aim is to include as many OPAF systems that fit the mentioned criteria, regardless of the date of publication
Screening	Reading title and abstract first, body if necessary later	The broad search string, combined with the mentioned databases, also delivers many false positives, which could easily be excluded by just looking at the title or abstract
Inclusion criteria	(I1) Automated feedback systems to improve public speaking skills (I2) Completeness of internal processing steps	Only OPAF systems are included, which collect sensor data and provide feedback, with the purpose of improving presentation skills
Exclusion criteria	(E1) Lack of pedagogical goal (E2) No implementation (E3) Not mentioned in scientific literature	‘Paper systems’ (as long as no prototype is developed, it is questionable if they are technically feasible) or commercial systems (which lack in terms of transparency) are excluded, as well as systems that are not mentioned in the scientific literature

Table 3 Search results for OPAFs based on inclusion and exclusion criteria, listed by number of matches in scientific databases

	ACM	*IEEE*	*MDPI*	*Science direct*	*Springer*	*Total*
Search location	No restrict.	No restrict.	Title + abstract	Title + abstract + keywords	Filter for computer science	
Retrieved publications	694	84	115	70	624	1587
Screening based on Inclusion and exclusion criteria	9	7	2	1	4	23
Snowball references						+22
Total						45

Source: Ortega-Arranz (2020, p.106)

3.3 DESMET methodology and evaluation process

The OPAF systems were evaluated using the DESMET methodology developed by Kitchenham (1998). This approach was chosen for its ability to compare and evaluate software tools systematically through feature analysis. We applied the screening mode, which allows for an efficient initial evaluation based on system documentation.

One limitation of the screening mode is its reliance on the subjective judgment of evaluators, which can introduce bias [Ortega-Arranz, (2020), p.108]. To mitigate this risk and keep the workload for individual evaluators low, we involved multiple evaluators in the process. Involving several evaluators also reduces the bias of pre-knowledge when evaluating multiple OPAFs. In total, 15 researchers (12 male and 3 female) who have a background in educational technologies, evaluated 14 different OPAF systems, each assessed using the scoring sheet developed in Section 3.1, which contains 83 functional features and 12 other relevant aspects. Three researchers independently reviewed each system, leading to a total of 3,990 individual evaluations.

Table 4 Search results regarding found OPAFs and a short description of these systems

	Discovered OPAF systems
S1	Presentation Sensei (2007) (Kurihara et al., 2007) is an automated presentation training system that uses speech and image processing to analyse a speaker's delivery during rehearsals. It provides real-time and post-presentation feedback on key aspects such as speaking rate, eye contact, filler words, and timing, helping users refine their presentation skills by reducing ineffective behaviours
S2	AwareMe (2016) (Bubel et al., 2016) is a public speaking training device that integrates cognitive behavioural therapy (CBT) principles to help users manage anxiety and improve speech delivery. It provides real-time feedback on voice pitch, filler words, and speaking rate through a wearable wristband with haptic and visual cues, increasing awareness of anxiety-related speech patterns to improve presentation skills
S3	The system developed by Dermody et al. (2015) is a real-time feedback tool that utilises Microsoft Kinect to analyse a speaker's body pose, facial expressions, and voice during presentations. It provides visual, text, and icon-based feedback on gaze direction, gestures, vocal tonality, dysfluencies, and speaking rate, allowing users to refine their presentation skills through both live and post-performance evaluations
S4	Quantle (2019) (Dermody and Sutherland, 2015) is a mobile presentation coaching app that provides real-time, privacy-preserving feedback on a speaker's vocal delivery. It analyses speech rate, pitch, pause duration, and readability complexity locally on the smartphone without transmitting data, helping users efficiently refine their speaking style
S5	Cicero V1 (2013) (Batrinca et al., 2013) is a multimodal virtual audience platform designed to provide public speaking training by analysing a speaker's nonverbal behaviours, such as gestures, tone of voice, and facial expressions. It uses automated behavioural descriptors and machine learning techniques to approximate expert judgments and provide detailed feedback on a user's performance
S6	Cicero VR (2019) (Chen and Fragomeni, 2019) is a virtual reality-based public speaking training tool that immerses users in realistic business presentation scenarios while measuring speech volume, speed, gesture, and eye contact. The system incorporates game-based learning mechanics that provide a structured training experience through various scenarios, including boardroom presentations and investor pitches, with real-time and post-session feedback

Table 4 Search results regarding found OPAFs and a short description of these systems (continued)

	Discovered OPAF systems
S7	NUSMSP (2015) (Gan et al., 2015) is a multi-sensor self-quantification framework for analysing presentation skills using static cameras, Kinect sensors, and wearable sensors (Google Glass). The system evaluates vocal behaviour, body language, engagement, and presentation state. It provides automated feedback based on multimodal sensor data to help presenters refine their delivery
S8	Presentation trainer V1 (2015) (Schneider et al., 2015) is a multimodal public speaking training tool that provides real-time feedback on nonverbal communication, including body posture, gestures, voice volume, pauses, and phonetic pauses. It uses Microsoft Kinect to track a speaker's movements and voice. It also provides corrective and interruptive feedback through visual and haptic cues to help users improve their presentation skills
S9	Presentation trainer VR (2019) (Schneider and Drachsler, 2019) is an immersive virtual reality-based public speaking training tool that extends the original presentation trainer (Schneider et al., 2015) by integrating a real-time VR feedback module. The system provides non-verbal communication feedback on posture, gestures, voice volume and pauses while allowing users to practice in a virtual classroom environment for a more immersive learning experience
S10	PresentMate (2015) (Lui et al., 2015) is a mobile application for self-regulated oral presentation training that uses a smartphone's built-in accelerometer and microphone to provide instant and post-session feedback on voice level, timing, and body movement. The system allows presenters to practice anytime, anywhere by providing electronic cue cards, slide viewing, and self-assessment reports, helping users refine their presentation skills without the need for an audience
S11	RAP (2018) (Ochoa et al., 2018) is a low-cost system designed to provide automated feedback on oral presentation skills using multimodal analysis from a simple camera and microphone setup. The system evaluates key aspects such as posture, gaze, voice volume, filled pauses, and slide quality. It generates feedback reports with multimodal recordings to help entry-level students improve their presentation skills
S12	Rhema (2015) (Ochoa et al., 2018) is a real-time, wearable public speaking assistant that uses Google Glass to provide live feedback on speech volume and rate. The system captures audio, processes it on a remote server, and displays feedback at short intervals to help users adjust their voice modulation during live presentations
S13	RoboCOP (2017) (Trinh et al., 2017) is an anthropomorphic robotic head designed as an automated coach for oral presentation rehearsals, providing voice feedback on speech quality, content coverage, and audience orientation. The system enhances the rehearsal experience by simulating an interactive and motivating audience, improving presentation quality through real-time coaching informed by academic mentors and validated by user studies
S14	Logue (2015) (Damian et al., 2015) is a real-time public speaking feedback system that uses a head-mounted display (HMD) and social signal processing techniques to analyse and provide instant feedback on nonverbal behaviours such as speaking rate, body energy, and openness. By providing unobtrusive, real-time behavioural cues, the system helps speakers dynamically adjust their delivery, improving their performance and self-awareness during presentations

After completing the reviews, the scores were aggregated, and a consensus-based approach was used to resolve any discrepancies. We achieved a consensus rate of 76%, with strong dissent occurring in only 0.7% of cases, where none of the three reviewers shared the same decision. In cases of strong disagreement, we re-examined our decisions, also considering the comments left in the comments section. This allowed us to clear up any misunderstandings that had already arisen. This evaluation process enabled us to accurately assess the state of current OPAF systems, identify missing features, and suggest improvements to enhance user acceptance and facilitate broader adoption.

4 Results

This section presents the results of our systematic analysis of 14 OPAFs (see Table 4), focusing on their feature coverage and evaluation in existing research.

Our results indicate a low average implementation rate of only 16% across all 83 identified functional features, highlighting significant gaps in functionality. Even excluding the organisation and central message category, the implementation rate remains at only 19%. However, it is worth mentioning that this aspect has not yet been a primary focus of OPAFs. These figures suggest that the observed OPAFs lack essential capabilities like the ones described in Section 4.1, which may contribute to their limited adoption.

To provide a structured analysis, the results are presented in two main sections. The first section examines feature coverage, categorising implemented functionality across different domains such as presentation delivery, feedback mechanisms, guidance, and expert collaboration. This analysis identifies both commonly implemented features and significant feature gaps, particularly in areas such as verbal-nonverbal congruency and presentation content organisation.

The second section focuses on the evaluation of OPAF systems in the literature. It examines the methodologies used to evaluate these systems and the specific criteria applied, such as usability, user experience, learning gains, and accuracy. Our findings reveal an imbalance in evaluation priorities, with a strong emphasis on usability and interaction quality, while learning gains and long-term pedagogical benefits remain underexplored. Furthermore, most evaluations have been conducted in controlled laboratory environments, raising concerns about the applicability of findings to real-world presentation settings.

Overall, these results highlight the fragmented development of OPAFs and the need for more comprehensive feature integration and standardised evaluation approaches. Addressing these gaps could enhance the effectiveness of these systems and increase their adoption in educational settings.

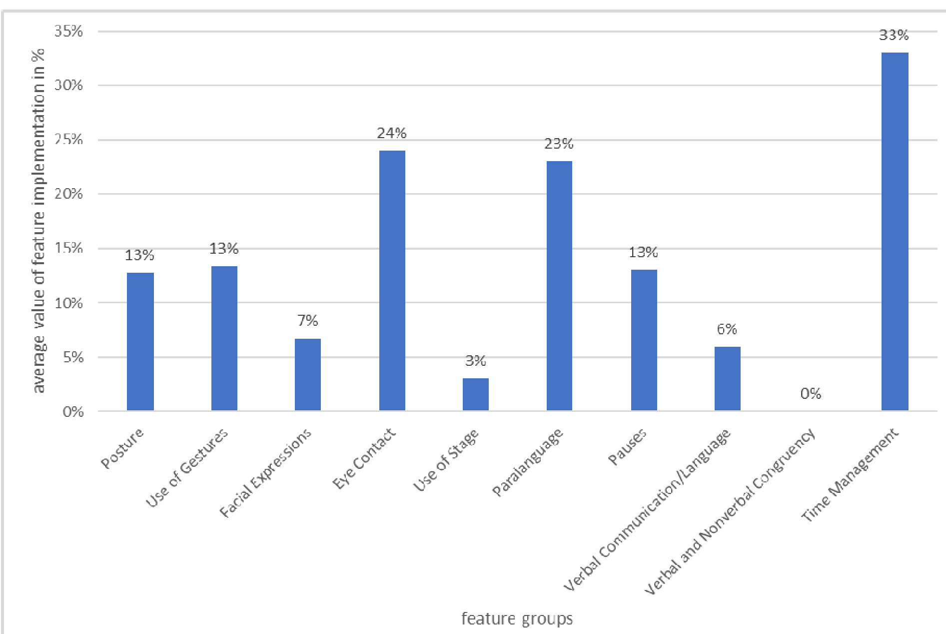
4.1 Feature coverage analysis and outstanding systems

After analysing the scoring sheets (see chapter 3.1) with each feature, we decided to create new subcategories for each main category (see chapter 2). We clustered these 41 features regarding the delivery section (see chapter 2) into 10 categories based on the work of Schneider et al. (2017). For example, features such as monitoring whether the user gives their back to the audience and whether the user's feet are firmly positioned between shoulder and waist width were grouped under ‘posture’. Similarly, features like monitoring if the user's voice volume is adequate and whether their voice is monotone were grouped under ‘paralanguage’. After categorising the features, the implementation levels were assessed by assigning a score of 2 for fully implemented features and 1 for partially implemented ones. Then, these scores were used to calculate the percentage of implementation for each feature group (see Figure 1). The analysis showed that features related to time management (33%), eye contact (24%), and use of voice (23%) were the most commonly implemented. However, none of the systems implemented any feature analysing verbal and nonverbal congruency. Overall, the average implementation level for features concerning the delivery of the presentation was 13%. The most implemented features included monitoring the use of filler sounds such as ‘ahm’ or ‘hmm’ (60%), assessing whether the user's voice volume was adequate (57%), monitoring whether the user's voice was monotone (53%), and identifying if the user paused while speaking (53%). Nevertheless, 15 out of 41 features in this category were not implemented at all.

In contrast, none of the analysed OPAFs provided any features addressing the 14 aspects relevant to the organisation and central message of a presentation. For example, features such as monitoring whether the introduction is designed to gain the attention and interest of the audience, ensuring that the main points are fully supported, and verifying whether a clear summary of the points discussed is provided were completely absent.

For features concerning supporting material, three relevant aspects were considered: monitoring whether supporting material provides insights into the topic, assessing whether visual aids are of high quality, and monitoring if the user is proficient in using the supporting material. Among the analysed OPAFs, S13 included features addressing two of these aspects, specifically monitoring whether supporting material provides insights into the topic and whether the user is proficient in using the material. Additionally, S11 included a feature for monitoring whether visual aids are of high quality. However, the remaining systems did not include any features related to supporting material.

Figure 1 Average feature implementation values regarding all 14 systems in terms of the main category delivery of the presentation based on the rubric of Schneider et al. (2017)



Feedback features were present in all analysed OPAFs, with varying levels of implementation. Based on the complexity dimension of feedback, simple verification was the most implemented type (57%), followed by correct response feedback (43%) and elaborated feedback explaining the reasons for the feedback (17%). Regarding the metacognitive dimension of feedback, 12 out of 14 systems provided at least one relevant feature. The most commonly implemented metacognitive feedback features were those helping users judge their performance (60%), become aware of their learning progress (43%), and evaluate how well they have learned (40%). Feedback timing also varied across systems, with 67% providing immediate feedback, 73% offering delayed feedback, and 43% incorporating both types. Examining the motivational aspect of feedback, 13 out of 14 OPAFs aimed to provide motivational features. These most commonly focused on the user's performance outcomes (73%), while fewer systems included features designed to trigger positive user emotions (20%) or support continuous improvement (13%).

Regarding guidance, three features were identified in this category. Among these, two features were fully implemented by S9: The presence of an onboarding process or tutorial to help users learn how to benefit from the system and the inclusion of examples of good practices for a presentation. The onboarding process was partially implemented by S6 and S12, while examples of good practices were partially implemented by S11 and S12. Additionally, S11 was the only OPAF to partially implement examples of bad practices.

Finally, collaboration between teachers and learners was limited across the analysed systems. All 14 OPAFs functioned as stand-alone applications, with six systems generating reports that, in theory, could be shared with teachers. However, none of the systems allowed teachers to adjust the system for personalised training, highlighting a lack of collaboration capabilities.

4.2 Evaluation of the published OPAF systems

The evaluation of the 14 OPAF systems covered six distinct categories (see chapter 2.7): Usability, technology acceptance, user experience, accuracy and performance, learning gains and social effects (e.g., presentation anxiety). While the evaluation categories were consistent, a total of 29 evaluations were made, indicating that some systems were evaluated under multiple criteria. For instance, 10 systems (34.48%) were assessed for user experience, and 7 systems (24.13%) were evaluated for usability. Technology acceptance and accuracy and performance each accounted for 4 systems (13.79%), social effects (6.89%), and learning gains for only 2 systems (6.89%) (see Figure 2).

This overlap in evaluation criteria highlights two key patterns. First, some systems were evaluated on multidimensional goals. For instance, S9 and S13, were assessed for their influence on usability and social effects, while S8 and S15 were evaluated for their impact on user experience and learning gains (see Table 4). Second, the distribution of classifications indicates variability in research priorities across studies. While user experience and usability were the most frequently evaluated criteria, learning gains and social effects were comparatively underrepresented. This discrepancy suggests that most studies emphasised short-term usability and interaction quality rather than long-term pedagogical outcomes.

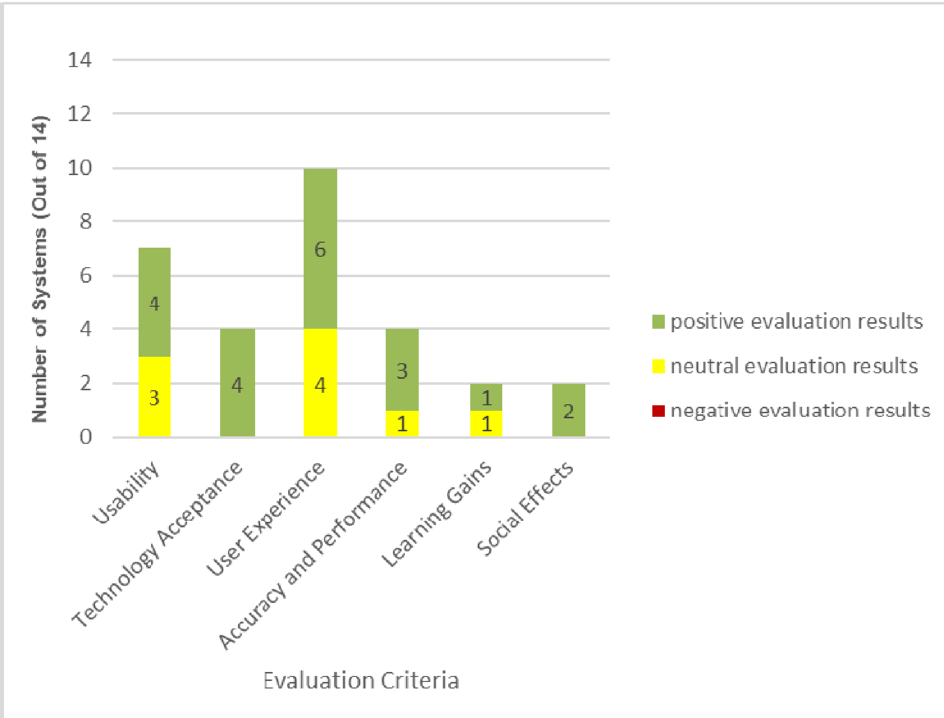
After taking a look on which criteria (see Figure 2) these systems have been evaluated, it is also important to figure out how these systems have been evaluated – therefore, Table 5 is providing an overview:

Table 5 Overview of the evaluation criteria regarding the different OPAFs

System	*Setting*	*Criteria*	*Methodology*	*No. participants/ samples*
S1	Lab-based	Accuracy and performance	Non-standardised questionnaire + interviews	3 participants
S2	Lab-based	Usability + user experience	Think aloud protocol	5 participants
S3	Lab-based	-	Survey	-
S4	Lab-based	Accuracy and performance	Comparative analysis	1017 samples
S5	Lab-based	Accuracy and performance	Comparative analysis	14 participants
S6	Lab-based	Usability + user experience	Non-standardised questionnaire	36 participants
S7	Lab-based	Usability + technology acceptance + user experience	Self generated evaluation rubric + comparative analysis	51 samples
S8	Lab-based	User experience + learning gains	Quasi-experimental research	40 participants
S9	Lab-based	Usability + technology acceptance + user experience + social effects	Non-standardised questionnaire	24 participants
S10	Lab-based	User experience	Non-standardised survey	20 participants
S11	Lab-based	Usability + technology acceptance + user experience + accuracy and performance	The system was reviewed by experts with a scoring rubric + non-standardised user experience questionnaire + interviews	3 participants to fill out scoring rubric, 83 participants to fill out the questionnaire and 9 participants for the interviews
S12	Lab-based	Usability + user experience	Brainstorming discussions + focus group interviews	30 participants
S13	Lab-based	Usability + technology acceptance + user experience + social effects	Non-standardised questionnaire and interviews	30 samples 12 participants
S14	Lab-based and in the wild	User experience + learning gains	Non-standardised questionnaire	First study: 15 participants Second study: 3 participants

The evaluation of some systems under multiple criteria suggests a lack of standardised methodologies for assessing OPAF systems. Researchers have adopted diverse tools and approaches to tackle the same criteria. For example, S2 employs the think-aloud method to assess usability, S7 utilises a self-generated evaluation rubric, S12 relies on focus group interviews, and S13 uses a self-generated questionnaire with a 7-point Likert scale (see Table 4). These variations result in discrepancies in how systems are evaluated.

Figure 2 System evaluation results based on their criteria (see online version for colours)



Another important point is the environment in which the experiments are carried out. Almost all (with only one exception) user evaluations were done in a lab-based setting instead of doing user evaluations in the wild.

5 Discussion

This study examined the current state of OPAF systems, focusing on both their implemented features (RQ1) and the evaluation methodologies used to assess them (RQ2). Our findings reveal a landscape of notable advancements, but also persistent gaps that hinder widespread adoption and effectiveness. While some OPAFs demonstrate strengths in specific feature categories, overall implementation remains fragmented, with only 16% of identified features being covered. Critical aspects such as the alignment of verbal and nonverbal communication, adaptive feedback mechanisms, and speaker movement tracking remain unaddressed. Furthermore, most systems rely on a rigid, rule-based feedback structure. This limits their ability to provide meaningful, goal-oriented learning experiences. The evaluation of these systems also presents challenges. Learning gains and social effects, which are crucial to assessing the educational impact of OPAFs, remain underexplored. Additionally, the lack of standardised assessment tools and the predominance of laboratory-based experiments restrict the applicability of findings to real-world settings. The contrast between positive user feedback and limited pedagogical effectiveness further underscores the need for more comprehensive evaluations that consider both usability and instructional value. To address these limitations, future research should focus on refining system functionalities, incorporating expert-driven feedback loops, and ensuring that evaluation methodologies reflect real-world applicability (Ochoa, 2022). The following sections discuss these findings in greater detail, outlining the strengths and limitations of existing OPAFs and providing recommendations for future development and research.

5.1 Features and their implementation status (RQ1)

The systems analysed were typically designed for specific purposes and excel in particular categories or subcategories. For example, the nonverbal communication category shows specialised areas: in posture, S8, S9, and S15 each cover 3 out of 6 features (50%). In facial expression, S8 implements 2 out of 3 features (66.66%), while S3, S6, and S11 achieve similar implementation rates in eye contact (66.66%). S5 leads in paralanguage with 4 out of 8 features fully implemented (50%). In the area of support materials, S13 covers 2 out of 3 features (66.66%), complemented by S11, which addresses the missing feature. Finally, for feedback timing, S1, S4, S5, S8, S10, and S14 stand out, providing both immediate and delayed feedback (100%).

However, these examples of excellence are contrasted by the broader picture, which reveals significant shortcomings in the implementation of functional features. Across all systems and features analysed, the overall implementation rate is 16%, underscoring the fragmented and incomplete nature of these solutions. Many features remain entirely unaddressed, limiting the systems' capacity to fully support users in developing public speaking skills. For example, none of the systems analysed provide feedback on spoken content, such as grammar or word choice, which are fundamental to effective communication. Similarly, no system evaluates stage usage, such as the speaker's positioning or movement patterns, which are critical to staying in contact with the audience. Most notably, none of the systems integrate verbal and nonverbal feedback to assess their alignment – a key determinant of authenticity and persuasiveness in presentations. Our interviewed experts consistently highlight the importance of congruence between verbal and nonverbal communication in building trust and credibility as also mentioned by Schneider et al. (2017), making this gap particularly striking.

The method by which feedback is delivered also reveals a consistent limitation across systems. All OPAFs analysed adopt a linear feedback approach: data about the speaker is captured and processed through pre-defined rules, which results in on-screen or audio feedback messages. While this approach provides corrective, metacognitive, or motivational feedback, it does not inform users about their progress toward achieving specific learning goals, nor does it offer recommendations for next steps or exercises tailored to their needs. This lack of adaptive and goal-oriented feedback limits the practical utility of these systems for continuous skill development.

A related issue is the insufficient attention to user guidance. Only S9 includes effective onboarding processes and examples of best practices, despite their proven efficacy in enhancing user engagement and learning effectiveness, as mentioned by Santos et al. (2024). Without clear guidance, users may struggle to interpret and act on the feedback provided, further diminishing the systems' impact. This issue is compounded by the tendency of most OPAFs to function as standalone tools, which limits their integration into educational contexts. Although six systems (S1, S3, S4, S6, S11, and S12) allow data sharing with educators, none provide the option for teachers to adapt feedback to individual student needs. This limitation raises concerns about the contextual reliability of automated feedback, as noted by expert interviews, and highlights the importance of enabling collaborative workflows where educators can supplement automated insights with their expertise.

The findings of this analysis underline a central paradox: while individual OPAFs demonstrate excellence in certain areas, their narrow focus and fragmented implementation prevent them from effectively addressing the broader spectrum of user needs. Addressing these limitations requires a strategic rethinking of OPAF design.

5.2 Evaluation of OPAF systems: criteria, limitations, and future directions (RQ2)

The evaluation of OPAFs reveals critical insights into their effectiveness and broader potential for adoption, but significant limitations remain. Learning gains and social effects are notably underrepresented in current evaluations, with only 2 out of 29 evaluations. This limited focus on these critical aspects undermines the understanding of how OPAFs contribute to the development of public speaking skills or influence factors like reducing public speaking anxiety and fostering confidence. The lack of emphasis on these dimensions leaves a significant gap in assessing the systems' holistic impact on users and their effectiveness as educational tools. Without sufficient evidence in these areas, the potential of OPAFs to deliver meaningful and lasting pedagogical benefits remains uncertain.

Another notable finding is the contradiction between the positive feedback from end users and the low level of feature coverage from the perspective of presentation experts. While students, as the primary users, value the ease of use and accessibility of OPAFs, experts emphasise the limited pedagogical usefulness of these systems, as only 16% of the features are implemented on average. This highlights a trade-off between usability and functionality. Students are likely to be motivated by a system that is easy to use, but if its usefulness remains too low, long-term engagement is unlikely. Therefore, researchers or developers need to provide ways to increase their usefulness, such as incorporating feedback from human experts to address complex or nuanced aspects of public speaking skills.

The lack of standardised assessment tools further complicates the evaluation of OPAFs. Most studies rely on customised questionnaires tailored to their specific research objectives, making direct comparisons between systems virtually impossible. While it is reasonable to include context-specific questions, standardised instruments such as the system usability scale for usability or the technology acceptance model for technology acceptance could facilitate a more consistent evaluation process. The inconsistency in methodologies limits the generation of actionable insights that could guide future development.

The dominance of laboratory-based experiments also raises questions about the practicality and scalability of OPAFs. 13 of 14 systems have been evaluated in controlled environments that do not reflect real-world conditions, such as varying acoustics or audience dynamics. While these studies provide valuable initial insights, their findings may not be fully applicable to the diverse environments in which users may engage with these systems. In addition, field experiments are constrained by issues such as limited hardware compatibility, high technical requirements, and accessibility barriers, particularly for systems that incorporate virtual reality capabilities. Addressing these challenges is critical to ensure that OPAFs are both scalable and practical for widespread use (Ochoa, 2022).

Future research should focus on expanding evaluation criteria to include learning gains and social impact, to ensure that these systems support sustained skill development and user confidence (Organisation for Economic Co-operation and Development (OECD), 2019). It is also important to actively seek out and publish studies on negative outcomes, as these provide critical insights for iterative system improvement. Standardised evaluation methodologies should be adopted to facilitate comparability across studies, allowing identification of best practices and common challenges. In addition, evaluations should extend beyond controlled laboratory environments to include field experiments to validate the practicality of these systems in real-world scenarios. (Ochoa, 2022). Finally, the integration of expert feedback mechanisms alongside automated feedback could improve the contextual relevance and overall accuracy of OPAFs, thereby addressing the limitations highlighted by the experts.

5.3 Future directions for OPAF development

To transition from research prototypes to widely adopted educational tools, several critical areas must be addressed:

1. Comprehensive feature integration: future systems should integrate more functional features from existing tools into a unified platform or adopt a modular approach that allows users to combine specialised functionalities based on their needs.
2. Adaptive and personalised feedback: OPAFs should incorporate adaptive learning mechanisms that track progress, align with user-specific learning goals, and offer tailored recommendations for improvement.
3. Enhanced user guidance: effective onboarding processes and best-practice examples should be standard features to improve accessibility and engagement.
4. Collaboration with educators: enabling customisable feedback workflows that allow teachers to refine and contextualise automated feedback will increase system adoption in educational settings.
5. Expanded evaluation criteria: future studies should prioritise learning gains and social impact to better understand the long-term benefits of OPAFs. Additionally, negative findings should be reported to facilitate iterative system improvements.
6. Real-world testing: more field-based experiments are needed to validate the practicality and scalability of OPAFs beyond controlled laboratory environments.

5.4 Limitations of the study

Before concluding, we acknowledge some limitations of this study. The study was primarily influenced by experts, most of whom have pedagogical backgrounds. As a result, their perspective on system features was primarily influenced by the perceived usefulness of covering as many aspects of public speaking skills as possible. This approach did not fully address the trade-offs involved in implementing certain features, such as technical feasibility or impact on usability. Additionally, this study focused on feature implementation rather than the effectiveness of individual features, which is an important aspect for future research.

Another limitation is the possible presence of publication bias in the studies reviewed. Of the studies reviewed, two-thirds report positive results, while the remaining one-third report neutral results. This imbalance limits the ability to critically evaluate weaknesses in system design. Negative evaluations would be particularly valuable for identifying ineffective features, understanding potential drawbacks of feedback mechanisms, and avoiding unproductive design approaches. The lack of such critical insights hinders the iterative improvement of OPAF systems.

In addition, many evaluations in the literature reviewed take a broad but shallow approach, attempting to address multiple objectives within a single study. While this provides a general overview, it often results in fragmented and superficial findings that lack the depth needed to draw precise conclusions. A more focused approach on specific aspects, such as the effectiveness of particular feedback features, would provide more actionable recommendations for system developers and educators.

Despite these limitations, this study highlights key areas for improvement and underscores the need for more comprehensive, unbiased, and methodologically rigorous evaluations of OPAF systems to facilitate their adoption and effectiveness in real-world educational settings.

6 Conclusions

The purpose of this study was to systematically analyse OPAFs (until November 2023) by evaluating their implemented features, identifying gaps, and assessing their effectiveness based on existing research. Through a structured methodological approach, we have provided an in-depth understanding of the strengths and limitations of OPAFs in improving public speaking skills.

To accomplish this, we first identified key features relevant to the development of public speaking skills through a combination of expert interviews and supporting literature. This led to the creation of a scoring sheet that included 83 functional features and 12 additional aspects crucial for effective automated feedback. Then, we conducted a systematic literature search of major academic databases to identify relevant OPAF systems. A total of 14 systems were selected based on predefined inclusion and exclusion criteria to ensure that only empirically validated tools were considered. These systems were evaluated three times (to achieve a better consensus) using the DESMET methodology in screening mode, with 15 researchers independently rating each system based on our scoring rubric.

Our results revealed significant gaps in feature implementation, with an average coverage rate of only 16% across all identified features. This low average can be explained by the fact that these systems were developed as research prototypes with a specific goal in mind. Other constraints, such as budget and manpower, may be additional reasons why trade-offs in feature implementation occurred. Key aspects such as the alignment between verbal and nonverbal communication, adaptive feedback mechanisms, and comprehensive guidance remain largely unaddressed. While certain systems demonstrated strengths in specific domains, none provided a holistic solution for public speaking training. In particular, feedback mechanisms were often rigid and lacked personalisation, limiting their ability to support continuous improvement. Additionally, most systems functioned as standalone tools without integration options for expert-driven feedback, reducing their applicability in educational settings.

The evaluation of OPAFs in existing research also revealed several shortcomings. The vast majority of studies prioritised usability and user experience, while only a small fraction assessed learning gains or social effects, such as the reduction of public speaking anxiety. In addition, a lack of standardised evaluation methods and an over-reliance on laboratory-based studies limited the applicability of findings to real-world scenarios. These limitations underscore the need for more comprehensive evaluations that address both usability and pedagogical effectiveness.

Our findings allowed us to identify the critical challenges to broader adoption and suggest actionable ways to improve the design and user experience of these systems. In doing so, we aim to contribute to the development of more effective tools for improving public speaking skills in educational settings. Future research should prioritise the development of adaptive and personalised feedback mechanisms, and adopt standardised evaluation protocols. In addition, real-world testing environments should be utilised to validate system effectiveness beyond controlled laboratory settings.

Ultimately, this study provides a baseline analysis that can inform the next generation of OPAF systems. Addressing the gaps identified will be critical to ensuring that these tools evolve from research prototypes into widely adopted solutions capable of meaningfully improving public speaking training. By fostering collaboration among educators, researchers, and system developers, the field can move toward more effective and scalable OPAFs that better meet the needs of learners and institutions.

Acknowledgements

This research was supported by the German Federal Ministry of Education and Research (BMBF) under grant number “01IS22075A” within the project “HyTea”.

Declarations

Grammarly and ChatGPT were used exclusively to improve language clarity and readability. The conception, design, analysis, and interpretation of the research were carried out entirely by the author.

All authors declare that they have no conflicts of interest.

All participants involved in the study provided informed consent prior to participation. Written consent was obtained from each participant.

References

Batrinca, L., Stratou, G., Shapiro, A., Morency, L-P. and Scherer, S. (2013) 'Cicero – towards a multimodal virtual audience platform for public speaking training', in Aylett, R., Krenn, B., Pelachaud, C. and Shimodaira, H. (Eds.): *Intelligent Virtual Agents*, in Lecture Notes in Computer Science, Vol. 8108, Springer Berlin Heidelberg, Berlin, Heidelberg, pp.116–128, DOI: 10.1007/978-3-642-40415-3_10.

Biocca, F., Harms, C. and Burgoon, J.K. (2003) 'Toward a more robust theory and measure of social presence: review and suggested criteria', P*resence Teleoperators Virtual Environ.*, October, Vol. 12, No. 5, pp.456–480, DOI: 10.1162/105474603322761270.

Blikstein, P. and Worsley, M. (2016) 'Multimodal learning analytics and education data mining: using computational technologies to measure complex learning tasks', *J. Learn. Anal.*, September, Vol. 3, No. 2, pp.220–238, DOI: 10.18608/jla.2016.32.11.

Brooke, J. (1996) 'SUS – a quick and dirty usability scale', *Usability Evaluation in Industry*, Vol. 189, pp.4–7.

Bubel, M., Jiang, R., Lee, C.H., Shi, W. and Tse, A. (2016) 'AwareMe: addressing fear of public speech through awareness', in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ACM, San Jose California USA, May, pp.68–73, DOI: 10.1145/2851581.2890633.

Chan, V. (2011) 'Teaching oral communication in undergraduate science: are we doing enough and doing it right?', *J. Learn. Des.*, February, Vol. 4, No. 3, pp.71–79, DOI: 10.5204/jld.v4i3.82.

Chen, J.Y.C. and Fragomeni, G. Eds. (2019) 'Virtual, augmented and mixed reality', *Applications and Case Studies: 11th International Conference*, *VAMR 2019*, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, Proceedings, Part II, in Lecture Notes in Computer Science, Vol. 11575, Springer International Publishing, Cham, DOI: 10.1007/978-3-030-21565-1.

Collins, J. (2004) 'Education techniques for lifelong learning: giving a powerpoint presentation: the art of communicating effectively', *RadioGraphics*, July, Vol. 24, No. 4, pp.1185–1192, DOI: 10.1148/rg.244035179.

Damian, I., Tan, C.S., (Sean) Baur, T., Schöning, J., Luyten, K. and 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*, ACM, Seoul Republic of Korea, April, pp.565–574, DOI: 10.1145/2702123.2702314.

Dermody, F. and Sutherland, A. (2015) 'A multimodal system for public speaking with real time feedback', in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ACM, Seattle Washington USA, November, pp.369–370, DOI: 10.1145/2818346.2823295.

Di Mitri, D., Schneider, J., Specht, M. and Drachsler, H. (2018) 'From signals to knowledge: a conceptual model for multimodal learning analytics', *J. Comput. Assist. Learn.*, Vol. 34, No. 4, pp.338–349, DOI: 10.1111/jcal.12288.

Edwards, C., Edwards, A., Spence, P.R. and Lin, X. (2018) 'I, teacher: using artificial intelligence (AI) and social robots in communication and instruction', *Commun. Educ.*, October, Vol. 67, No. 4, pp.473–480, DOI: 10.1080/03634523.2018.1502459.

Gan, T., Wong, Y., Mandal, B., Chandrasekhar, V. and Kankanhalli, M.S. (2015) 'Multi-sensor self-quantification of presentations', in *Proceedings of the 23rd ACM International Conference on Multimedia*, ACM, Brisbane Australia, October, pp.601–610, DOI: 10.1145/2733373.2806252.

Gillis, R.L. and Nilsen, E.S. (2017) 'Consistency between verbal and non-verbal affective cues: a clue to speaker credibility', *Cogn. Emot.*, May, Vol. 31, No. 4, pp.645–656, DOI: 10.1080/02699931.2016.1147422.

Hattie, J. and Timperley, H. (2007) 'The power of feedback', *Rev. Educ. Res.*, March, Vol. 77, No. 1, pp.81–112, DOI: 10.3102/003465430298487.

Heinicke, M.R., Juanico, J.F., Valentino, A.L. and Sellers, T.P. (2022) 'Improving behaviour analysts' public speaking: recommendations from expert interviews', *Behav. Anal. Pract.*, March, Vol. 15, No. 1, pp.203–218, DOI: 10.1007/s40617-020-00538-4.

Hülsmann, F., Kopp, S. and Botsch, M. (2017) 'Automatic error analysis of human motor performance for interactive coaching in virtual reality', 26 September, arXiv: arXiv:1709. 09131, DOI: 10.48550/arXiv.1709.09131.

ISO (2018) *Ergonomics of Human-System Interaction – Part 11: Usability: Definitions and Concepts*, ISO 9241-11:2018 [online] https://www.iso.org/standard/63500.html (accessed 17 December 2024).

ISO (2019) *Ergonomics of Human-System Interaction – Part 210: Human-Centred Design for Interactive Systems* [online] https://www.iso.org/standard/77520.html (accessed 19 December 2024).

Keller, J.M. (1987) 'Development and use of the ARCS model of instructional design', *J. Instr. Dev.*, September, Vol. 10, No. 3, pp.2–10, DOI: 10.1007/BF02905780.

Kerby, D. and Romine, J. (2009) 'Develop oral presentation skills through accounting curriculum design and course-embedded assessment', *J. Educ. Bus.*, December, Vol. 85, No. 3, pp.172–179, DOI: 10.1080/08832320903252389.

Kitchenham, B. (1998) 'Evaluating software engineering methods and tools part 12: evaluating DESMET', *ACM SIGSOFT Software Engineering Notes*, Vol. 23, No. 5, pp.21–24, https://doi.org/10.1145/290249.290255.

Kitchenham, B., Linkman, S. and Law, D. (1997) 'DESMET: a methodology for evaluating software engineering methods and tools', *Computing & Control Engineering Journal*, Vol. 8, No. 3, pp.120–126.

Kurihara, K., Goto, M., Ogata, J., Matsusaka, Y. and Igarashi, T. (2007) 'Presentation sensei: a presentation training system using speech and image processing', in *Proceedings of the 9th International Conference on Multimodal Interfaces*, ACM, Nagoya Aichi Japan, November, pp.358–365, DOI: 10.1145/1322192.1322256.

Laugwitz, B., Held, T. and Schrepp, M. (2008) 'Construction and evaluation of a user experience questionnaire', in Holzinger, (Ed.): *HCI and Usability for Education and Work*, *A. in Lecture Notes in Computer Science*, Vol. 5298, pp.63–76, Springer Berlin Heidelberg, Berlin, Heidelberg, DOI: 10.1007/978-3-540-89350-9_6.

Levasseur, D.G., Dean, K.W. and Pfaff, J. (2004) 'Speech pedagogy beyond the basics: a study of instructional methods in the advanced public speaking course', *Commun. Educ.*, July, Vol. 53, No. 3, pp.234–252, DOI: 10.1080/0363452042000265170.

Lui, A.K-F., Ng, S-C. and Wong, W-W. (2015) 'a novel mobile application for training oral presentation delivery skills', Lam, J., Ng, K.K., Cheung, S.K.S., Wong, T.L., Li, K.C. and Wang, F.L. (Eds.): in *Technology in Education. Technology-Mediated Proactive Learning*, *in Communications in Computer and Information Science*, Vol. 559, pp.79–89, Springer Berlin Heidelberg, Berlin, Heidelberg, DOI: 10.1007/978-3-662-48978-9_8.

McGrath, C.H., Guerin, B., Harte, E., Frearson, M. and Manville, C. (2015) 'Learning gain in higher education', in *RAND Corporation eBooks*, https://doi.org/10.7249/rr996.

Monteiro, D. et al. (2024) 'Effects of audience familiarity on anxiety in a virtual reality public speaking training tool', *Univers. Access Inf. Soc.*, March, Vol. 23, No. 1, pp.23–34, DOI: 10.1007/s10209-023-00985-0.

Morreale, S.P., Myers, S.A., Backlund, P.M. and Simonds, C.J. (2016) 'Study IX of the basic communication course at two- and four-year U.S. Colleges and Universities: a re-examination of our discipline's 'front porch'', *Commun. Educ.*, Vol. 65, No. 3, pp.338–355, DOI: 10.1080/03634523.2015.1073339.

Mory, E.H. (2004) 'Feedback research revisited', in Jonassen, D.H. (Ed.): *Handbook of Research on Educational Communications and Technology*, 2nd ed., pp.745–783, Lawrence Erlbaum Associates Publishers.

Ochoa, X. (2022) 'Multimodal systems for automated oral presentation feedback: a comparative analysis', in Giannakos, M., Spikol, D., Di Mitri, D., Sharma, K. and Hammad, R. (Eds.): *The Multimodal Learning Analytics Handbook*, Springer International Publishing, Cham, DOI: 10.1007/978-3-031-08076-0.

Ochoa, X. and Zhao, H. (2024) 'OpenOPAF: an open source multimodal system for automated feedback for oral presentations', *J. Learn. Anal.*, November, pp.1–25, DOI: 10.18608/jla.2024.8411.

Ochoa, X., Domínguez, F., Guamán, B., Maya, R., Falcones, G. and Castells, J. (2018) 'The RAP system: automatic feedback of oral presentation skills using multimodal analysis and low-cost sensors', in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, ACM, Sydney New South Wales Australia, March, pp.360–364, DOI: 10.1145/3170358.3170406.

Organisation for Economic Co-operation and Development (OECD) (2019) *Skills for 2030: Concept Note* [online] https://www.oecd.org/content/dam/oecd/en/about/projects/edu/education-2040/concept-notes/Skills_for_2030_concept_note.pdf (accessed 10 February 2025).

Ortega-Arranz, A. (2020) *Supporting Practitioners in the Gamification of MOOCs Through Reward-Based Strategies*, Doctoral Thesis, Universidad de Valladolid, Spain.

Patil, M., Patil, V. and Katre, U. (2024) 'Unspoken science: exploring the significance of body language in science and academia', *Eur. Heart J.*, January, Vol. 45, No. 4, pp.250–252, DOI: 10.1093/eurheartj/ehad598.

Rodríguez-Triana, M.J., Prieto, L.P., Martínez-Monés, A., Asensio-Pérez, J.I. and Dimitriadis, Y. (2018) 'The teacher in the loop: customizing multimodal learning analytics for blended learning', in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, ACM, Sydney New South Wales Australia, March, pp.417–426, DOI: 10.1145/3170358.3170364.

Rubenson, D. (2021) 'Good presentation skills benefit careers – and science', *Nature*, May, Vol. 594, No. 7864, pp.S51–S52, DOI: 10.1038/d41586-021-01281-8.

Santos, I., Felizardo, K.R., Gerosa, M.A. and Steinmacher, I. (2024) *Software Solutions for Newcomers' Onboarding in Software Projects: A Systematic Literature Review*, August 28, 2024 [online] http://arxiv.org/abs/2408.15989 (accessed 29 August 2024)

Saukh, O. and Maag, B. (2019) 'Quantle: fair and honest presentation coach in your pocket', in *Proceedings of the 18th International Conference on Information Processing in Sensor Networks*, ACM, Montreal Quebec Canada, April, pp.253–264, DOI: 10.1145/3302506.3310405.

Schneider, J., Börner, D., van Rosmalen, P. and Specht, M. (2015) 'Presentation trainer, your public speaking multimodal coach', in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ACM, Seattle Washington USA, November, pp.539–546, DOI: 10.1145/2818346.2830603.

Schneider, J., Börner, D., van Rosmalen, P. and Specht, M. (2017) 'Presentation trainer: what experts and computers can tell about your nonverbal communication: presentation trainer', *J. Comput. Assist. Learn.*, April, Vol. 33, No. 2, pp.164–177, DOI: 10.1111/jcal.12175.

Schneider, R. and Drachsler, (2019) 'Beyond reality – extending a presentation trainer with an immersive VR module', *Sensors*, Vol. 19, No. 16, p.3457, August, DOI: 10.3390/s19163457.

Schreiber, L.M., Paul, G.D. and Shibley, L.R. (2012) 'The development and test of the public speaking competence rubric', *Commun. Educ.*, July, Vol. 61, No. 3, pp.205–233, DOI: 10.1080/03634523.2012.670709.

Sharma, K. and Giannakos, M. (2020) 'Multimodal data capabilities for learning: What can multimodal data tell us about learning?', *Br. J. Educ. Technol.*, September, Vol. 51, No. 5, pp.1450–1484, DOI: 10.1111/bjet.12993.

Svendsen, Ø. (2022) 'Theorizing public performances for international negotiations', *Int. Stud. Q.*, July, Vol. 66, No. 3, p.sqac031, DOI: 10.1093/isq/sqac031.

Tanveer, M.I., Lin, E. and (Ehsan) 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*, ACM, Atlanta Georgia USA, March, pp.286–295, DOI: 10.1145/2678025.2701386.

Trinh, H., Asadi, R., Edge, D. and Bickmore, T. (2017) 'RoboCOP: a robotic coach for oral presentations', *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, June, Vol. 1, No. 2, pp.1–24, DOI: 10.1145/3090092.

Venkatesh, V. and Bala, H. (2008) 'Technology acceptance model 3 and a research agenda on interventions', *Decis. Sci.*, May, Vol. 39, No. 2, pp.273–315, DOI: 10.1111/j.1540-5915.2008.00192.x.

Venkatesh, V. and Davis, F.D. (2000) 'A theoretical extension of the technology acceptance model: four longitudinal field studies', *Manag. Sci.*, February, Vol. 46, No. 2, pp.186–204, DOI: 10.1287/mnsc.46.2.186.11926.

Worsley, M. (2018) 'Multimodal learning analytics' past, present, and potential futures', in *Companion Proceedings of the Eighth International Conference on Learning Analytics and Knowledge (LAK 2018)*, Sydney, Australia, SoLAR, 5–9 March, [online] http://bit.ly/lak18-companion-proceedings.

Zimmerman, B.J. and Schunk, D.H. Eds. (1989) 'Self-regulated learning and academic achievement', in *Springer Series in Cognitive Development*, Springer New York, New York, NY, DOI: 10.1007/978-1-4612-3618-4.