forum acusticum 2026
8-12 September · Graz, Austria

THE AUDIENCE REACTOR: AN EMBODIED SPARRING PARTNER FOR PUBLIC-SPEAKING TRAINING

Idun Elvira Christensen[1], Vivien Johanna Wegener[1], Ivan Varbanov[1], Ferenc József Marton[1], Ïo Valls-Ratés[1], Oliver Niebuhr[1,2]

[1] *University of Southern Denmark, Sønderborg, Denmark*
[2] *AllGoodSpeakers Aps, C/O Domicilet, Ellegårdvej 36, 6400 Sønderborg*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

Public-speaking skills comprise a set of interrelated abilities, including vocal delivery, linguistic appropriateness, and audience-oriented communication. However, this multidimensionality makes it challenging to provide individualized, high-quality feedback, particularly within time-constrained curricula and in contexts where instructors lack specialized training in oral communication pedagogy.

To address this problem, the Audience Reactor was developed. The Audience Reactor is a humanoid-like robot that gives feedback during the user's speech, allowing them to actively improve and see their performance based on the robot's reaction. The reactions include facial expressions, body colour changes, motion, and sound cues and icons indicating the speech parameters most important to improving the overall speech quality. The robot is designed to give the impression of actively listening to the user, with the feedback being playful to encourage its use. The algorithm from which the feedback is derived is a speech analysis algorithm provided by the company AllGoodSpeakers. A web application allows the user to see their progress and a more detailed speech analysis through the saved results from previous sessions.

Keywords: *public speaking training, speech feedback, voice analysis, humanoid robot*

1. Introduction

Public speaking training has been delivered through traditional, digital, and experimental approaches [1], with webinars and seminars remaining widely used, while online courses continue to gain popularity.

Existing products can be broadly categorised into three main groups: human-guided training, software-based solutions, and embodied system developments. Technological advancements offer access to low-cost, flexible digital and virtual training solutions, which increasingly compete with traditional formats.

Companies such as VirtualSpeech [2] and OvationVR [3] use virtual reality to simulate public speaking scenarios and measure performance metrics like eye contact, filler words or speaking pace. AI-driven coaching apps like Orai [4] or Yoodli [5] focus on speech quality analysis through smartphones. There are also professional in-person or online sessions available, as well as recorded speech tutorials on different digital platforms. These solutions demonstrate that digital feedback is popular and feasible.

However, there are several limitations of current approaches, as most available solutions lack physical interaction, which reduces the sense of presence typically associated with speaking in front of an audience[6], [7], [8] . Furthermore, a study published in The British and Irish Orthoptic Journal [9] shows that after 15 minutes of VR use, participants aged 19-25 had signs of temporary eye fatigue. As a result, longer sessions could be more energy-draining for users, especially those with vision problems.

2. Why

Research demonstrates the feasibility of an anthropomorphic robot as a successful speech training aid [10], and commercial products already provide digital feedback. Currently, no widely adopted solution combines an audience reaction robot with a website platform that includes overall user performance statistics and is accessible outside of a laboratory.

The Audience Reactor aims to fill this gap. By introducing a "physical presence" and embodied feedback, the Audience Reactor offers a tangible, fun, and engaging alternative for practising public speaking that bridges the gap between purely digital feedback and human presence.

The Audience Reactor can best be defined as a niche,

***Corresponding author**: Idun.e.Christansen@gmail.com.*

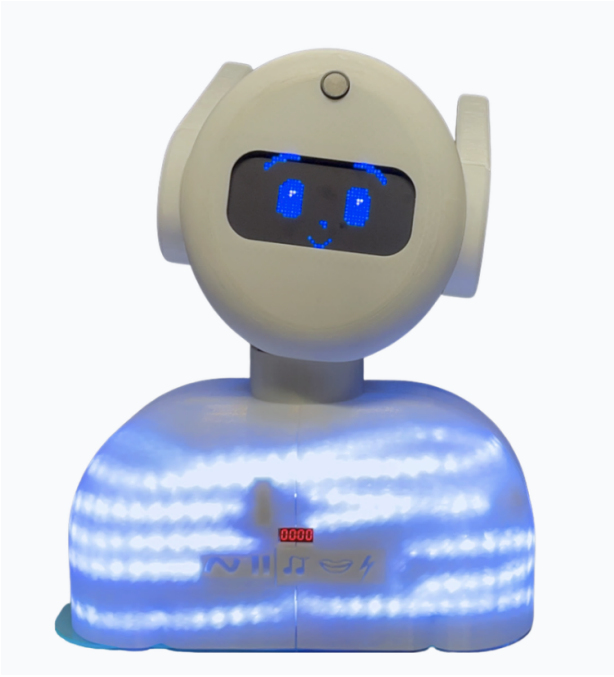

Figure 1. Audience Reactor Idle Mode.

complementary solution within the broader public speaking training market. Interviews conducted with industry experts, start-up accelerators and a pedagogical consultant have confirmed that the device addresses a real problem and the vocal parameters are crucial components of a good presentation presence. “Content can be bad, but if you’re a really good speaker, you can get away with it”(K.H.Thomsen, personal communication, November 2025).

3. Implementation

The Audience Reactor is a humanoid upper-body robot placed on a table to act as an audience. It analyses the user’s speech via a microphone, evaluating key parameters with a speech analysis algorithm. Based on these ratings, it provides visual feedback through facial expressions, body colour, and icons approximately every 30 seconds, allowing users to adjust their performance in real time. After the session, results, progress tracking, and improvement tips are available via a website. The fully assembled system is shown in Figure 1.

3.1. Microphone

Because the head follows the speaker as the speaker moves around during the session, a directional microphone can be used. In this case, a hyper-cardioid polar pattern was chosen. The microphone is a dynamic moving coil microphone. The microphone is supported by an amplifier, as the microphone itself is not able to add enough gain to produce a recording that can be analysed by the speech analysis algorithm. An amplifier with a 48 dB gain is used. The amplifier is externally powered to avoid noise introduction to the recording, which happened when powering the

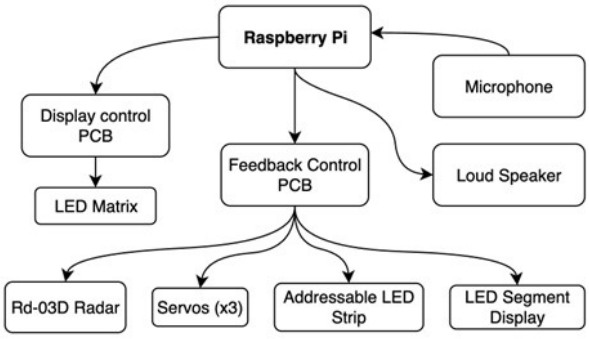

Figure 2. Hardware System Architecture Block Diagram.

microphone from the Raspberry Pi. As the amplifier contains a phantom power option, the microphone could be replaced by a shotgun microphone to achieve even more noise rejection and a clearer recording of the speech. This was not done in this case due to time constraints.

The microphone has a flat frequency response to achieve a good vocal analysis, as using a microphone with a shaped frequency response would bias the recording and, therefore, the speech analysis. The same principle is applied to the choice of amplifier. The final setup was tested, and the feedback algorithm variables were tuned by experts based on their experience. However, these values can be tuned to achieve a stricter or more relaxed feedback.

3.2. Software

A local PostgreSQL database was used to store system data. The database has tables for storing web sessions, user-relevant data, and the categorized values from the coordinator script. The coordinator script contains the code that communicates the categorized values to the different parts of the feedback system. For security considerations, all sensitive data is stored encrypted using Argon2 encryption, such as user passwords.

The priorities regarding the firmware were modularity of code and non-blocking of other functions. This was done by dividing the different functions into “tasks” that were then called in the main function, which acted as a coordinator function.

3.3. Electronics

The system is controlled by multiple PCBs and a central Raspberry Pi coordinating overall behaviour. Each PCB manages its own subsystem, enabling modularity, reducing processing load, and improving maintainability. The feedback PCB connects via USB and controls the MCU through text commands, while UART receives speaker position data from a radar module. GPIO pins control a segment display, an LED strip, and three servos.

The addressable LED strip requires a higher logic level for data communication. To ensure reliable signal transmission and prevent damage to the microcontroller, the PCB incorporates a logic level shifter and a series current-limiting resistor.

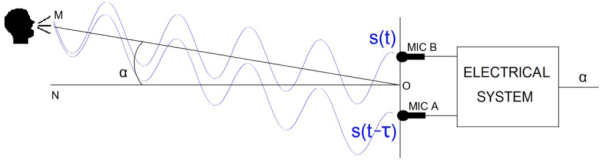

Figure 3. Schematic illustration of the radar.

The face display requires sixteen connections to be controlled, meaning a separate development module with a custom breakout board was created, forming the Display control PCB.

3.4. Design

The Audience Reactor's appearance went through multiple iterations, each one getting feedback from potential users. The Audience Reactor's characteristic pointed ears come from wanting the robot to seem more friendly and to allow the user more freedom in how the Audience Reactor looks, as they are modular and can be switched out easily. The pointed ears also strengthen the perception that the Audience Reactor is listening attentively to the user.
The torso is designed as four clip-together quadrants, and the head as two screw-fastened parts, allowing for 3D printing on standard printers and straightforward assembly. Both structures include connection points for components.

3.4.1. Movement

The head movement system of the Audience Reactor was designed to support expressive but simple, human-like non-verbal feedback while remaining mechanically robust. The head movement was decomposed into two mechanically independent subsystems:

- Head rotation for tracking the speaker's position
- Head nodding for when the speaker is improving

The whole rotation mechanism sits in the torso under the neck and consists of one main bearing, which sits around the neck, helping with radial forces and smooth motion, and four smaller bearings under the neck to help with the axial forces. The mechanism is controlled by a servo that the neck rests on.
A radar sensor is used to detect the speaker's position. After initialization, the radar calculates the speaker's position as an angle, which is sent to a servo motor responsible for rotating the Audience Reactor's head. The radar task manages serial communication with the radar, processing incoming data via a finite state machine and forwarding calculated angles to a servo task when tracking is active (Figure 3).
Inspired by the neck mechanism of the humanoid robot Adam [8], the nodding system uses a compact linkage design adapted for the Audience Reactor, incorporating a ball joint to provide two additional degrees of freedom (Figure 4).
The "servo task" in the code is in charge of controlling the servos, which turns the head and initiates

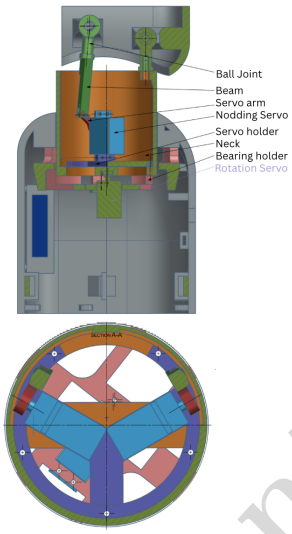

Figure 4. Cutaway graphics of the movement assembly.

the nodding, while making sure that the servo can move without blocking other parts of the code. It takes 30 milliseconds for the servo to move the head of the Audience Reactor to a new position. The nodding behaviour is implemented as a simple finite state machine that operates within the servo update function. Rather than doing a nod as a single action, which would block other parts of the code, it is instead broken down into several discrete states that are advanced over time as the servo completes each movement. This also adds more possibilities to control the nodding motion.
The final mechanical design prioritizes simplicity and robustness, which limits the range and speed of head motion. Rapid or large amplitude nodding and tilting motions were intentionally avoided to reduce wear on the joint linkages, structural components, and static component connections.

3.5. Feedback

The feedback consists mainly of visual, but also one audible element. The audible elements are incorporated only at the start and the end of the session. This creates a focus pull back to the Audience Reactor in case of distractions, but does not interrupt the session. To encourage and create a fun learning experience, all of the feedback elements are designed to be either gamified or have an encouraging nature.

3.5.1. Feedback Elements

The visual feedback consists of an LED matrix which displays facial expressions, LED's which illuminate the body of the Audience Reactor, LEDs which illuminate Icons, and a timer display.
Facial expressions are an important factor for emotional communication [11], hence it is important that the Audience Reactor can display facial expressions. To be able to do this, and still give the robot a characteristic look, a 64x32 LED matrix is used. The facial expressions are based on anime characters' facial ex-

pressions, following the concept of gamification, and also making the emotion of the robot very clear. Four facial expressions are implemented, detailed in the feedback algorithm section below.
An addressable LED strip is used to illuminate the body in different colours. Correlating to the facial expressions, the body colour of the Audience Reactor changes to underscore the meaning of the facial expressions. The body colours are red, green, and yellow, where green correlates to the happy face indicating the user to keep going, yellow indicating some problems, and red indicating serious issues. The LED strip is also used to illuminate icons in the torso of the body. The icons correspond to five speech parameters, selected with the aims to keep the speech-analysis algorithm computationally lean while still covering all aspects of charismatic speaker impact. Firstly, they cover all main prosodic dimensions: melody, timing, energy, and voice quality. Pitch variability, operationalized as the standard deviation of F0, indicates how expressive or monotonous a speaker sounds and captures melodic movement more robustly than simple pitch range [12]. Pause count represents timing, fluency, and rhythmic structuring. Pitch level indicates whether a voice sounds relatively high or low, adjusted according to the probable gender of the user, which is derived from average pitch level. Mean F1, the first resonance frequency, reflects jaw opening and articulatory expansion and serves as a cue to articulation clarity and rhythmic structuring ability [12]. Finally, H1-A3 is a robust spectral-slope measure related to voice quality, and perceived vocal and, together with mean F1, can be a proxy for voice energy (perceived loudness). Secondly, the 5 parameters were selected to effectively represent all three pillars of a speaker's charismatic impact: perceived competence, self-confidence, and passion [13]. Pitch level is particularly relevant to perceived passion, pause count and H1-A3 to perceived self-confidence, and mean F1 and pitch variability to perceived competence. During a session, the speech parameter that needs to be improved most urgently for the overall score of the user to improve is illuminated [14].
Over the speech parameter icons is a timer to display the time left in the speech to the user. The display used for the timer is a small 4-digit 7-segment display. As previously mentioned, the audible feedback is implemented to pull the focus back to the Audience Reactor in case of distractions. A speaker is implemented to play a sound at the start and at the end of the speech. The chosen sounds are playful. At the beginning, there is a race start beeping, and in the end, an applause sound is used to encourage the user.

3.5.2. Feedback Algorithm

The software architecture of the Audience Reactor system is modular, consisting of components that execute separate tasks under the control of an orchestrator script. The web application serves as the entry point; users log in to an account and initiate a session by specifying the speech duration. Upon initiation, the main coordinator script receives the duration via a REST API call and starts the audio recording script. It routes the recorded audio through the Machine Learning (ML) and categorizer components for analysis and handles communication between the Pi, the web application, and the PCB. During recording, calculated speech values from the most recent 40 seconds are sent to the PCB every 20 seconds. Once the recording concludes, the coordinator script transmits the full speech data back to the web application for storage in a PostgreSQL database and subsequent visualization. The audio recorder script manages recording via two concurrent threads: one records the full speech, while the second continuously saves the most recent 40 seconds to a buffer.
The feedback system is driven by the machine learning algorithm. The ML requires an audio recording of the speech, which is captured using the audio recording scripts. This recording is processed by the ML to calculate feedback parameters such as speech quality indicators. These five parameters are then mapped to four broader categories: Extremely Low, Average, Good, or Extremely High. The five examined parameters are weighted by a priority number (1 being lowest, 5 being highest) in the following order:

- F0std (pitch variability)
- pauseCount
- meanF0 (pitch level)
- meanF1 (first resonance frequency)
- H1minusA3 (energy)

At the same time, each parameter is classified into one of four performance levels: good, middle, extreme low, or extreme high. These levels also have a severity score from 1 (good) to four (extreme). A small deviation has a lower score, while a strong deviation has a higher score. To decide what feedback to show, the system multiplies the importance of a parameter by its severity. This creates a total score for each parameter. To determine the displayed icon, the categorizer identifies the parameter with the highest priority weighted value within the non-optimal categories and transmits this to the PCB to light up the corresponding icon.
The categorizer calculates a weighted sum for each of the four broader categories. The category with the highest impact determines the robot's state:

- Good: Green colour, smiling face.
- Average: Yellow colour, confused face.
- Extremely Low: Red colour, sleeping face.
- Extremely High: Red colour, overwhelmed face.

The extremely low mode correlates to a low pitch variability, and overall "boring" speech delivery, while extreme high mode correlates to a very agitated and overexcited speech delivery. When the analysis of the session is finished, the robot returns to its idle state, which consists of a white LED colour for the body, as it is the most neutral, all of the icon LEDs are

Figure 5. Robot states depicted in five feedback modes.

shut off, and the face expression returns to the smiling state. The different states are shown in Figure 5. To encourage the user, a nod is initiated every time the state of the robot improves.

3.6. User Journey

To start practising with the Audience Reactor the user has to first connect their device to the local network hosted on the Raspberry Pi named "AGS". Then they have to go to a browser and visit the website of the Audience Reactor, by typing in "ags.local" on their device. On the website, the next step is to create an account or log in to an existing one. After logging in, the user can set the duration of the speech and initiate a session by pressing the start button. Then, a 10-second countdown starts, and the duration is displayed on the timer on the robot. During that the user can get in front of the Audience Reactor and prepare for the speech. After the 10-second preparation time expires, a beeping sound indicates the start of the session, the timer on the robot starts, and the user can start the speech. Every 30-40 seconds, feedback is given in the form of the combined facial expression and body colour and illumination of the icon of the parameter with the most influence in the past 30-40 seconds. When the speech ends, the timer stops, and an applause sound is played, indicating the recording is over. When the result for the whole speech is calculated and sent to the website, the robot goes back to idle mode. This indicates that the results are ready for the user on the website, so the next step is to go back to the website, press the "View Results" button, and the list of all the results is shown. Here, the user can select any speech result from the past, as well as the current one. When the user chooses a result from the list, a radar chart is displayed with the scores of the 5 different speech parameters representing the whole session, as shown in Figure 6. In addition, general tips on how to improve the speech are provided. There is an option to export the session results in JSON format. The user can start another session by going back to the "Start Session" tab on the website.

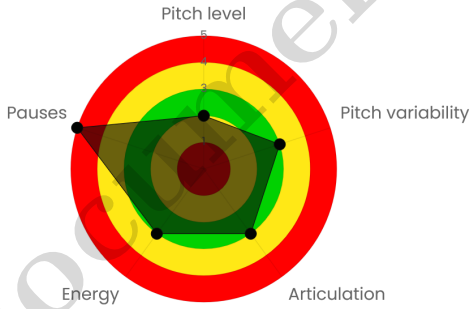

Figure 6. Example of a results chart showing the values of the five vocal parameters after the speaker's presentation. The green circle depicts the optimal value for each of the parameters.

4. Discussion and Conclusion

4.1. Proof of Concept Achieved

The Audience Reactor project successfully manages to bring to reality the futuristic idea of practising a speech in front of a robot, which actively gives feedback during sessions. Through student feedback over the course of the research phase, as well as from university events after the project's completion, we can see that the design of the robot, the feedback from the lights, and the website, are clear, understandable, easy to navigate, and spark the curiosity in students, who represent one of the primary target group.

At university events, the Audience Reactor was met with enthusiasm and curiosity. People, especially students, were intrigued and, if given the option, would be willing to practice with the Audience Reactor. The test sessions conducted received positive feedback. Especially emphasizing the ability to see the parameters needing improvement in real time and having the overview at the end as more detailed feedback, in addition to the ability to train without the fear of judgment, and working on overcoming anxiety related to public speaking.

4.2. Future Work & System Optimization

Even though the Audience Reactor is still a prototype, there are still ways to improve the finished product in the future. New iterations could focus on improving audio fidelity by replacing the current hyper-cardioid microphone with a shotgun microphone to improve noise rejection. Developing an extensive and robust testing suite for the integration of the machine learning components will benefit software security. Additionally, head tilting and nodding are features that would improve the overall user experience with the robot. If subsequent development phases are to take place, they will prioritise even greater comprehensive user testing with target demographics. UI improvements will include a dark mode and potentially colour-coded illumination for the speech parameter icons.

The Audience Reactor is designed to be used primarily in a solo training setting or as an assistance to traditional public speaking training. In solo settings, it allows students to practice, get feedback, and improve their public speaking skills without people present, reducing anxiety and allowing easier access to training possibilities. In traditional training, it can allow the trainer to focus on other aspects of practice, helping streamline and improve the training session.

5. Acknowledgments

This paper is written based on a report for a university course on this topic, which was prepared collaboratively with Juan Diego Lynggaard, Nikita Satsuk, Wiktoria Lasowa, Stanislaw Basiuka and Zuzanna Siwinska. We would like to thank our project partners for their valuable contributions and teamwork throughout this project. In addition, we would like to express our gratitude to our university supervisors Shouvik Chaudhuri and William Greenbank, for their academic guidance, constructive feedback and support during this project. We would like to thank the industry experts, startup accelerators, and pedagogical consultants for their participation in 1-to-1 interviews.

References

[1] DataIntelo, *Public speaking training market report*, 2025.

[2] VirtualSpeech, *Virtualspeech – ai-powered soft skills training in vr and online*, 2025.

[3] Ovation VR, *Ovation vr – home page*, 2025.

[4] Orai, *Orai – home page*, 2025.

[5] Yoodli, *Yoodli – home page*, 2025.

[6] D. Verdonik, P. Rupnik, and N. Ljubešić, "Disfluencies in public and private speech," *Language and Speech*, 2025.

[7] O. Niebuhr and J. Michalsky, "Virtual reality simulations as a new tool for practicing presentations and refining public-speaking skills," in *Proc. of the 9th International Conference on Speech Prosody 2018*, Poznan, Poland, 2018, pp. 309–313.

[8] I. Valls-Ratés, O. Niebuhr, and P. Prieto, "Unguided virtual-reality training can enhance the oral presentation skills of high-school students," *Frontiers in Communication*, vol. 7, p. 910 952, 2022.

[9] S. Banstola, K. Hanna, and A. O'Connor, "Changes to visual parameters following virtual reality gameplay," *The British and Irish Orthoptic Journal*, vol. 18, no. 1, pp. 57–64, 2022.

[10] D. Forghani, *System Development and Evaluation of a Social Robot as a Public Speaking Coach*. Springer, 2025.

[11] L.-P. Morency, J. Smith, and W. Zhao, "Multimodal analysis of public speaking anxiety and vocal performance," *Journal of Voice*, vol. 40, pp. 123–135, 2026.

[12] O. Niebuhr and R. Skarnitzl, "Measuring a speaker's acoustic correlates of pitch–but which? a contrastive analysis based on perceived speaker charisma," in *Proc. of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, 2019, pp. 1–5.

[13] I. Poggi, "The "seeds" of charisma," *Multimodal Argumentation and Rhetoric in Media Genres*, vol. 14, p. 263, 2017.

[14] O. Niebuhr, "Akustisches charisma: Profiling für digitale rhetorik," *Akustik Journal*, vol. 20, no. 2, pp. 12–18, 2020.