Educational Administration: Theory and Practice
2024, 30(5), 15191 - 15199
ISSN: 2148-2403
https://kuey.net/

Research Article

Enhancing Public Speaking Skills Through AI-Powered Analysis And Feedback

Soham Padia[1*], Jainam Patel[2], Divyam Jain[3], Sweedle Machado[4], Stevina Correia[5], Monali Sankhe[6]

[1*]Dwarkadas J. Sanghvi, College of Engineering, Mumbai, Maharashtra, India sohampadia10@gmail.com
[2]Dwarkadas J. Sanghvi, College of Engineering, Mumbai, Maharashtra, India pateljainam25@gmail.com
[3]Dwarkadas J. Sanghvi, College of Engineering, Mumbai, Maharashtra, India divyamjain910@gmail.com
[4]Dwarkadas J. Sanghvi, College of Engineering, Mumbai, Maharashtra, India sweedle.machado@djsce.ac.in
5Dwarkadas J. Sanghvi, College of Engineering, Mumbai, Maharashtra, India stevina.dias@djsce.ac.in
[6]Dwarkadas J. Sanghvi, College of Engineering, Mumbai, Maharashtra, India monali.sankhe@djsce.ac.in

Citation: Soham Padia, et al (2024), Enhancing Public Speaking Skills Through AI-Powered Analysis And Feedback, *Educational Administration: Theory and Practice, 30(5),* 15191 - 15199
Doi: 10.53555/kuey.v30i5.8524

ARTICLE INFO

ABSTRACT

Public speaking anxiety, or glossophobia, affects a significant portion of the population, hindering effective communication. Existing solutions, such as traditional coaching and generic online courses, often fail to provide personalized, realtime feedback. These approaches lack the ability to dynamically analyze both verbal and non-verbal cues in a holistic manner. This paper presents an AI-driven application designed to enhance public speaking skills through personalized feedback on voice modulation, facial expressions, and speech content. By employing technologies such as SpaCy and NLTK for text processing, OpenCV and YOLO for facial expression and gesture recognition, and OpenAI-Whisper and SpeechRecognition for speech analysis, the system provides users with targeted, actionable insights to improve their performance. Experiments conducted using a custom dataset containing videos of speeches demonstrate an overall system accuracy of 87.73%, with individual component accuracies of 92.25% for text processing, 76.45% for facial and gesture recognition, and 92.5% for speech analysis.

Index Terms—Public Speaking, Artificial Intelligence, Natural Language Processing, Facial Expression Analysis, Voice Modulation, SpaCy, NLTK, OpenCV, YOLO, SpeechRecognition, OpenAI-Whisper

I. INTRODUCTION

Artificial intelligence (AI) is becoming a key part of our daily lives, improving fields like education, healthcare, and personal development. One important area where AI can help is in public speaking, often hindered by glossophobia (the fear of public speaking) [1]. Public speaking is essential for success in many areas, yet many people struggle with anxiety and lack effective training tools.

Research indicates that AI can significantly enhance public speaking training. For instance, those systems that evaluate non-verbal behavior provide interactive and adaptive training in real-time for trainees to get instant feedback on body language and facial expressions [2]. In addition, AI-based speech analysis tools for content and delivery provide immediate and detailed feedback, which was absent in traditional methods, hence enabling users to identify the errors quickly and polish their speaking skills [3]. These AI advances turn training in public speaking into an energetic and tailored practice that serves the needs of clients much more effectively than onesize-fits-all coaching or Internet courses. Conversely, current AI-based solutions frequently miss the incorporation of both verbal and nonverbal feedback in a comprehensive model and are thus severely hampered in their effectiveness for delivering a fully training-rich experience [4].

Online tools, such as Yoodli, have also made forays into the public speaking training arena with AI-driven feedback on various features in a user's performance [5]. While Yoodli is very useful, one of the main differentiators for our application is to provide more granular feedback on non-verbal cues and the integration of more features in analysis that will enhance this training experience, ensuring comprehensive assessment of both verbal and non-verbal skills in communication.

In order to combat these, this research work aims to develop an AI-powered application to provide users with personalized feedback about their public speaking. The user uploads his video/audio and the processing of

video and audio input is done for speech, facial expression, and gestures through OpenCV [6] and YOLO [7], while text processing is handled by SpaCy [8] and NLTK [9]. It uses OpenAI-Whisper [10] and SpeechRecognition [11] to provide tips and feedback on voice modulation, facial expression, and structure of content [4]. This enables the user to build confidence and communicational skills and, therefore, minimizes the daunting factor that comes with public speaking.

Experiments with the AI-powered app, conducted using a custom dataset containing videos of speeches, returned very promising results in enhancing public speaking skills. The users realized that they could talk with more confidence and improved delivery. They appreciated immediate feedback that is personalized about their performances. It has picked up most of the common public speaking mistakes—like filler words and inconsistent pacing—and given action enablers that are quite easy to act on. Feedback from such tests reveals high decreases in anxiety levels and a clear sense of improved communication effectiveness. The overall system accuracy from the experiments was 87.73%, while the accuracies from different components are 92.25% for text processing, 76.45% for facial and gesture recognition, and 92.5% for speech analysis [4]. These results indicate that the AI-driven approach could offer a valid excellent alternative to the existing traditional approaches of training in public speaking, with a much more tailored and responsive learning experience.

The development and the features of this AI application are discussed in the paper, including its technology and its implications for public speaking training. It is an AI-based application that will provide access to high-quality public speaking training for all people to continuously learn and improve. Current AI approaches and their effectiveness in enhancing communication skills are reviewed in the next sections in order to understand the context and impact of these innovations.

II. LITERATURE SURVEY

Public speaking is one of the essential competences needed in both the professional and personal world today. However, many people do not have the confidence and skill due to not being exposed enough; therefore, they develop phobias, such as glossophobia. Recently, especially with technological advancement in AI, there is a lot of potential help to solve these problems. This literature review identifies some of the key research papers using AI techniques, more specifically Deep Learning and Natural Language Processing, in order to provide better public speaking training.

Remarkably, among the proposed applications, it is stated that Deep Learning and NLP assist users in breaking communication barriers. The application focuses on developing the confidence and fluency of communication by facial emotion detection and vocabulary-grammar suggestion [12]. A fundamental approach to a more complex system provided the case for the incorporation of emotional intelligence into public speaking training. Following this, it introduces an automated system that can interpret human nonverbal behaviors in real time. These features include the Multimodal Affective Reactive Characters, which allow touching aspects of realism by capturing arm and posture movements, facial expressions, gaze behavior, and lip synchronization [13]. Since this particular system focuses more on the non-verbal cues, it complements the verbal feedback mechanisms to provide an overall communication skill training.

Deeper into the emotional aspects, the work of strives for the comprehension of the emotional states by speech. In this respect, different computational methodologies are examined, such as support vector machines and a Gaussian mixture vector autoregressive approach to classify negative emotions from positive emotional states using acoustic-prosodic features and domain-specific cues [14]. The paper thus identifies that emotional speech analysis is complex and requires a sophisticated model to accurately capture these nuances.

It describes how there has been an advance in speech processing and, thereby, an increasing tendency of deploying DNN for recognition and separation. This paper proposes a new architecture that bases its implementation on DNN in order to predict the features of speech with higher accuracy, thus proving the superiority of deep learning techniques over the others [15]. This proves that AI can understand and develop various features of speech rendition.

While current AI-enabled solutions have made tremendous headway in enhancing public speaking training, a few gaps still exist. Most existing systems today tend to pay attention to either the verbal or non-verbal cues and do not integrate both comprehensively. Besides, emotional analysis in speech is complex and requires more refined models. The paper tries to bridge these gaps by integrating a bouquet of technologies like SpaCy [8], NLTK [9], OpenCV [6], YOLO [7], OpenAI-Whisper [10], SpeechRecognition [11] in a holistic and integrated approach to public speaking training arena. This application is capable of presenting feedback on the verbal and non-verbal communication dimensions in real-time and in a highly personalized manner by the usage of these sophisticated techniques; this boosts up the confidence and proficiency levels of the users. Careful actionable feedback is guaranteed with the use of deep learning models, thus setting a new benchmark for AI-driven public speaking training tools [4].

III. METHODOLOGY

Effective public speaking requires a combination of strong verbal and non-verbal communication skills. To help individuals improve these skills, an AI-driven application has been developed in this research work that

provides personalized feedback on various aspects of public speaking. This section outlines the system architecture, key components, and detailed processes involved in analyzing and enhancing public speaking performance.

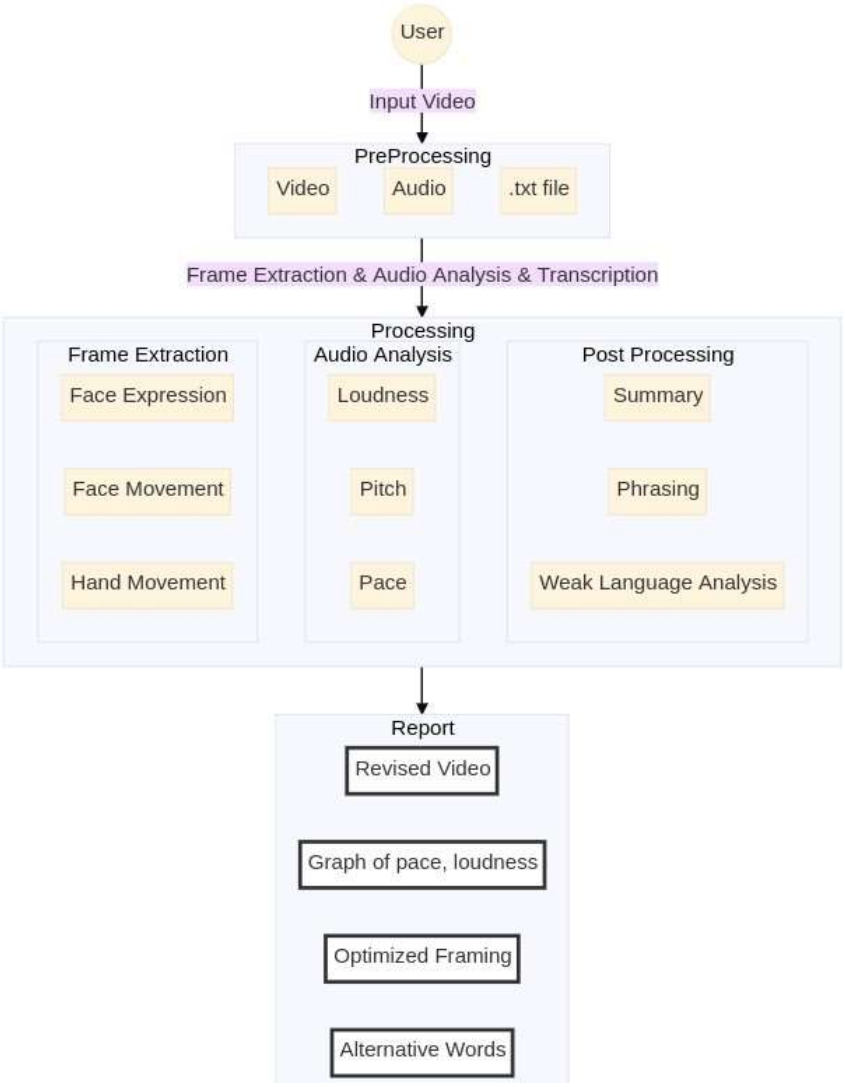

Fig. 1. Block Diagram of the System

The block diagram (Figure 1) illustrates the flow of the system from input to output. The system processes video, audio, and text data to generate a comprehensive feedback report for the user.

A. System Architecture

The AI-driven application integrates advanced technologies such as Natural Language Processing (NLP), Deep Learning, and Computer Vision to analyze voice modulation, facial expressions, and speech content [16]. The system is designed to assist users in overcoming common public speaking challenges, providing a supportive platform for users ranging from students to professionals seeking to refine their communication skills.

The architecture of the system comprises several key components: input (video format), segmentation (video, audio, and text), preprocessing, analysis, and output (feedback generation). Each component plays a critical role in ensuring the system delivers accurate and actionable feedback to the user.

B. Input (Video Format)

The system begins by collecting a video input from the user. This video is then divided into three segments: video, audio, and text. Each segment undergoes preprocessing to ensure it is suitable for detailed analysis.

C. Preprocessing

1) *Video Preprocessing:* The video segment is processed to extract frames using OpenCV. This step ensures that each frame is suitable for further analysis. Key frames are then extracted to focus on facial expressions and body movements.
2) *Audio Preprocessing:* The audio segment is normalized and segmented into frames to facilitate detailed analysis. This process prepares the audio data for subsequent analysis steps, including pitch, loudness, and pacing evaluations.

3) *Text Preprocessing:* The text segment is obtained by transcribing the audio using OpenAI-Whisper. The transcribed text is then tokenized and cleaned using SpaCy and NLTK to prepare it for further analysis.

D. Analysis

1) Video Analysis: There are many successive steps through which non-verbal communication cues in videos are assessed. The AI-driven application makes use of YOLO and OpenCV in processing and analyzing the video frames extracted earlier. The first step in this direction is the preprocessing step to extract the individual frames of the video, after which each is fine-tuned and prepared for analysis. Afterwards, these frames are fed into the YOLO model where various detections, including facial expressions, face movements, and hand movements, do take place. This detection is of essence in evaluating nonverbal communication skills. Basically, this YOLO model detects and classifies several body and facial gestures through the extraction of bounding boxes around these features. For example, it classifies whether the user has maintained eye contact, hand gestures, and facial expressions. OpenCV can be used to further enhance these detections using the video frames for better quality of capture.

These include duration and frequency of establishing eye contact, range, and variety of hand gestures, and expressiveness of facial movements. After the extraction of these metrics, an analysis follows to reach conclusions on the level of the user's non-verbal communication abilities. This indepth analysis will show the user exactly where improvements need to be made—for instance, in eye contact, using many more hand gestures, or facial expression to create the speech engaging. This ensures that users get focused feedback on their non-verbal communication, something very essential in public speaking.

2) Audio Analysis: Analysis of the audio includes speech pace, pitch, and loudness. Transcribe audio with Whisper transcription, and plot speaking rate over time with Python's matplotlib. The audio is first segmented into frames and then normalized before a detailed analysis is carried out. This details a graph, as shown in Fig. 2. The graph illustrates the modulations of speaking rate—words per second—throughout the speech. The plot highlights some of the moments where there are variations in speech velocity. The X-axis measures time in seconds, and the Y-axis refers to speaking rate. This graph could be useful in modulating a speaker's rhythm and pacing and is very practical in improving a speaker's delivery. Plotting this information on graphs enables users to instantly understand many of the important facets of their speaking performance that will help in making improvements. AI transcription and data visualization will further enhance such feedback by providing insightful input on the style of delivery of a speaker.

The dataset will further be used in matching peaks and ridges in the graph of a speaker's pace. This will enable identification at time pauses between words, occurring after a filler word. This additional analysis gives fine-grained insight into how filler words affect speech flow and allows giving specific feedback on reduction in filler words and fluency improvement.

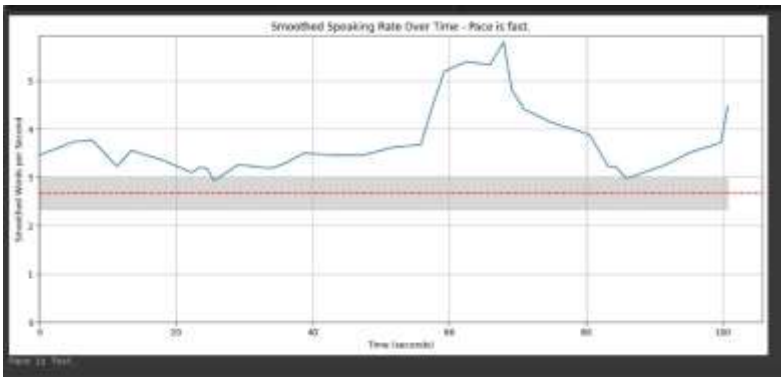

Fig. 2. Speech Pace

3) Text Analysis: The textual analysis component will then process and analyze the obtained text from the user's speech through SpaCy and NLTK. First, OpenAI-Whisper would transcribe the audio, converting the speech into text. This transcribed text is then tokenized and cleaned of stop words and punctuation to maintain only the relevant information for processing.

The AI-powered text summarization tool uses SpaCy to compute the word frequencies by discarding the stopwords and punctuation, and scoring sentences based on these normalized word frequencies. This system computes the number of sentences used to summarize as a function of video length to generate a summary with an optimum quantity of information. Summarization helps the user quickly get the main points of the long content, hence increasing manageability and ease of understanding. These key sentences identify the essence of the original content, providing a quick, insightful overview. Clearly visible, the summary demonstrates what the application is capable of regarding distilling crucial information—thus acting as an invaluable resource for quickly grasping complex texts. This detailed analysis aids in improving the structure and content language in speeches [17].

##ORIGINAL DOC
What's going on guys, Matt McNamara here with another video and in today's video I'm going to walk through my one minute sales pitch. So when I'm sitting down with a prospect, when I'm visiting with them, when I'm meeting with them, this is exactly what I say. It's really important that you're able to sell yourself and your business and how you might be able to help them out either now or down the road in 60 seconds or less. So here we go. Good morning, my name is Matt McNamara. I'm a business development manager with Formcraft and what I'd like you to do is take a moment to envision your dream office space. What do you want it to look like, feel like, and flow like? It's my job at Formcraft to make sure that every single tenant in Philadelphia knows that we can make your dream office a reality both on time and on budget, creating an environment where people love to come into work every single day and business thrives. So if you want an office space like Google, we'll Google it up, a conservative space with cubes for a call center, we got you covered there too. Now your journey with Formcraft starts right now and I get it, it can be a bit overwhelming, a project of this size and scope, but this is what we do every single day. We love to do it every single day. We're going to love to do it for you folks every single day. So what I'm going to do now, I'm going to pass it over to our design team and they're going to walk you guys through a fun activity that we have planned to get to know you better, what you guys are all about, and how we might be able to help you out with your future office space. So thanks for inviting us in. So that is what I say every time I am meeting with a prospect or in the future when I'm attending sales pitches with our design team. This is exactly how I'll introduce the meeting, introduce myself, and then pass it over to our designer. So that's the one minute pitch. If you guys are having trouble coming up with a one minute pitch, let me know. I'd love to help you guys out. Comment below what you guys think and like and subscribe. Thanks a lot.

SUMMARY
So what I'm going to do now, I'm going to pass it over to our design team and they're going to walk you guys through a fun activity that we have planned to get to know you better, what you guys are all about, and how we might be able to help you out with your future office space. It's my job at Formcraft to make sure that every single tenant in Philadelphia knows that we can make your dream office a reality both on time and on budget, creating an environment where people love to come into work every single day and business thrives. What's going on guys, Matt McNamara here with another video and in today's video I'm going to walk through my one minute sales pitch.

Fig. 3. Speech summary

E. Output (Feedback Generation)

The results from video, audio, and textual analyses are combined into a feedback report. Verbal and non-verbal performance in public speaking is given in detail and action-oriented to users. The feedback will be shown in an analysis tab, which includes detailed evaluation and recommendations to improve.

1) Text Analysis Output: As shown in Fig., the text analysis section summarizes, transcribes, identifies key takeaways, and provides suggestions for class content refinement. In a nutshell, it summarizes what the speech is all about and is auto-generated through the identification and extraction of the key sentences by SpaCy. One can also see the full transcript of the speech where one's words can be reviewed in detail. The key takeaways are those things that top off a speech page by giving the important elements a second whirl so that a user really gets the critical points they conveyed. It provides suggestions to bring out the speech better, including sentence rewording and avoiding weak language.

Fig. 4. Textual analysis output

2) Audio Analysis Output: The section on audio analysis includes graphs and metrics for pacing, pitch control, voice modulation rating, and articulation clarity. The pacing graph in Fig. 5 provides insight into how speaking rate is distributed across the time the speaker is talking. It helps to note and eliminate irregularities from the speech. Pitch control is represented with a circular graph, showing a minimum and maximum pitch. Voice Modulation Rating This aspect gives a qualitative value rating in regards to the user, where it is concerning the vocal dynamics and the areas for maintaining engaging volume levels. Articulation Clarity Over Time This shows a tracking over time of improvements or changes in speech clarity.

Fig. 5. Audio analysis output

3) Video Analysis Output: As shown in Fig., the video analysis section provides feedback on facial expression, body language, eye contact, and hand gestures. The system, through YOLO and OpenCV, will grade non-verbal cues to ascertain whether the speaker is engaging with the audience or not. Facial expression comments let a user know how he or she is expressing their emotions. It is used to analyze the inclusion of body gestures and posture, while eye contact feedback is used to ensure hold-up with the audience. Hand gestures are also assessed, providing information on their effectiveness and appropriateness during the speech.

Fig. 6. Video analysis output

This feedback report is systematically developed to help users improve their self-confidence while speaking in public by giving them relevant advice in both verbal and non-verbal communication of ideas. The system, through its advanced AI insights incorporated within, gives users relevant and precise recommendations that go a long way in upgrading their performances.

F. Algorithm/Pseudo Code for the Deep Learning Model

```
# AI-Driven Public Speaking Enhancement Application

# User Registration/Login
WHILE not logged in
        Display welcome, gather credentials, create profile, log in

# Speech Input and Feedback Loop WHILE actively speaking
FOR EACH input_type IN [voice, video, speech_content]
        Display interface, capture input
        IF input_type IS voice
                Preprocess voice (normalize, segment into frames)
                Extract features (MFCCs, pitch, intensity)
                Analyse voice data using Deep Learning model
                Generate voice feedback
        ELIF input_type IS video
```

Preprocess video (frame extraction, resizing)
Detect and analyze facial expressions
Generate video feedback on expressiveness
ELIF input_type IS speech_content
Preprocess text (tokenization, removing stopwords)
Analyze speech content (grammar, coherence, relevance) Generate content feedback
Display feedback
IF all feedbacks available
Generate and display comprehensive feedback

This research develops an AI application that enhances public speaking skills through real-time, personalized feedback.
Integrating voice and facial expression analysis with content optimization, the system offers targeted improvement suggestions, boosting users' confidence and proficiency. Continuous updates ensure alignment with the latest AI advancements. This comprehensive method evaluates all aspects of public speaking, providing a robust tool for communication skill enhancement [16].

IV. SYSTEM IMPLEMENTATION

Experimentation for this algorithm involved several tests and evaluations to assess its functionality, robustness, and performance.

A. Dataset Explanation

The evaluation of the proposed application utilizes two distinct datasets:
1) Public Speaking Video Dataset*:* For this research, a new dataset was created, consisting of nearly 800 videos, including public speaking videos fetched from websites like YouTube, Kaggle, and the UCI repository, and manually recorded public speaking sessions by our peers. Each video entry is elaborated with a URL video, speaker information and duration which makes the resource complete for training and testing our models.
2) Filler Word Dataset: There is a unique, custom-made dataset of approximately 1100 entries made by our team solely on the use of filler words in speech. It makes a difference between sentences where words like "like" are used as fillers, for example, "Like, I can't believe he said that," and the use of words such as "like" as non-fillers in "I like chocolate more than vanilla." This dataset is very critical to the training of our system on the identification and handling of filler words in speech; it highly refines speech delivery in public speaking. Each entry is supported by a sentence ID, filler word, its position in the sentence, context, and meaning, facilitating exact analysis and feedback.
This dataset supports AI feedback on filler word usage by providing context-specific examples, enhancing the model's ability to differentiate between filler and non-filler contexts.

B. Evaluation Measure

The AI-driven application is evaluated based on several performance parameters, including efficiency and accuracy.
1) Efficiency: Efficiency in the system is quantified by the processing speed of OpenAI Whisper when analyzing speech from video content. We measure efficiency in terms of the time required and the frame rate achieved during processing. The following table provides detailed metrics for both CPU and GPU processing units, reflecting the time efficiency and frames per second (FPS) managed for various video lengths.
These measurements are critical for understanding the system's capability to handle processing loads efficiently, ensuring that users experience minimal delay in receiving feedback on their public speaking performances. The significantly reduced processing times on the GPU demonstrate the benefits of hardware acceleration in AI-driven applications.

TABLE I EFFICIENCY OF SPEECH ANALYSIS

Processing Unit	Video Length (minutes)	Time Taken (seconds)	FPS
CPU	1	10	6
CPU	2	20	6
CPU	5	50	6
GPU	1	2	30
GPU	2	4	30
GPU	5	10	30

2) Accuracy: The precision of the feedback provided by the system is critically evaluated across three main dimensions: voice modulation, facial expression recognition, and speech content accuracy. The system achieves an accuracy of 92.5% in voice modulation, 76.45% in facial expression recognition, and 92.25% in text processing. This meticulous evaluation ensures that users receive highly accurate and actionable insights into their public speaking abilities, facilitating effective improvement.

TABLE II ACCURACY OF SYSTEM COMPONENTS

Component	Tools Used	Accuracy
Text Processing	SpaCy, NLTK	92.25%
Facial/Gesture Recognition	OpenCV, YOLO	76.45%
Speech Analysis	OpenAI-Whisper, SpeechRecognition	92.5%
Overall System		87.73%

3) Comparison Table: A table comparing the performance of the AI-driven application against traditional public speaking training methods highlights the improvements in user confidence and speaking proficiency.

TABLE III COMPARISON OF AI-DRIVEN APPLICATION VS. TRADITIONAL METHODS

Parameter	AI Application	Traditional Methods
Real-time Feedback	Yes	No
Personalized Feedback	Yes	Limited
Non-verbalCues Analysis	Yes	No
Speech Optimization	Yes	No
Scalability	High	Low

CONCLUSION

Public speaking challenges, exacerbated by anxiety and traditional training limitations, are addressed by our AI-driven application, which enhances skills through advanced Natural Language Processing and Deep Learning. The application provides personalized feedback on voice modulation, facial expressions, and speech content, demonstrating a significant improvement over conventional methods in our tests. These results underscore the application's effectiveness in boosting user confidence and communication abilities, offering a more engaging and accessible approach to mastering public speaking. This study confirms the potential of AI to revolutionize public speaking training, making it more effective and widely accessible.

FUTURE WORK

Looking ahead, this research will focus on several key development areas:

- Algorithm Optimization: Enhancing the precision and customization of feedback by refining the algorithms.
- Expanding Language Support: Adding multilingual support to serve a wider audience.
- Integration with Virtual Reality (VR): Implementing VR to create more engaging and realistic public speaking simulations.

These advancements will reinforce the application's role as a leading AI-driven educational tool, broadening access to effective public speaking training.

REFERENCES

[1] T. Pfister and P. Robinson, "Real-time recognition of affective states from nonverbal features of speech and its application for public speaking skill analysis," *IEEE Transactions on Affective Computing*, vol. 2, pp. 66–78, April 2011.

[2] S. N. Kirillov, V. T. Dmitriev, and S. O. Aleksenko, "Machine learning algorithms based on hidden markov models in low-speed speech codecs for assessing speech quality," in *2020 2nd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA)*, pp. 408–412, Nov 2020.

[3] A. F. Muhammad, D. Ekky Pratama, and A. Alimudin, "Development of web based application with speech recognition as english learning conversation training media," in *2019 International Electronics Symposium (IES)*, pp. 571–576, Sep. 2019.
[4] E. Kimani, T. Bickmore, H. Trinh, and P. Pedrelli, "You'll be great: Virtual agent-based cognitive restructuring to reduce public speaking anxiety," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 641–647, Sep. 2019.
[5] "Yoodli — Free Communication Coach — yoodli.ai." https://yoodli.ai/. [6] "Home — opencv.org." https://opencv.org/. [Accessed 25-07-2024].
[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2016.
[8] "spaCy · Industrial-strength Natural Language Processing in Python." [9] "NLTK :: Natural Language Toolkit — nltk.org." https://www.nltk.org/. [Accessed 25-07-2024].
[10] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," 2022.
[11] "GitHub - Uberi/speech_recognition: Speech recognition module for Python, supporting several engines and APIs, online and offline. — github.com." https://github.com/Uberi/speech_recognition. [Accessed 25-07-2024].
[12] S. V. Jadhav, S. R. Shinde, D. K. Dalal, T. M. Deshpande, A. S. Dhakne, and Y. M. Gaherwar, "Improve communication skills using ai," in *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*, pp. 1–5, March 2023.
[13] M. Hoque and R. W. Picard, "Automated coach to practice conversations," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 703–704, Sep. 2013.
[14] R. Ranjan, "Analysis of speech emotion recognition and detection using deep learning," in *2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, pp. 1–5, Dec 2022.
[15] M. S. Chauhan, R. Mishra, and M. I. Patel, "Speech recognition and separation system using deep learning," in *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, pp. 1–5, Sep. 2021.
[16] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech emotion recognition using deep learning techniques: A review," *IEEE Access*, vol. 7, pp. 117327–117345, 2019.
[17] M. Kaloev and G. Krastev, "Comparative analysis of activation functions used in the hidden layers of deep neural networks," in *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pp. 1–5, June 2021.