

V-EXAM – Voice-based Explanation Assessment Model

DR. D. BANUMATHY¹, K. DHIKSHANTH², S. AJAY³, C. ARUNKUMAR⁴

¹HOD, Department of Computer Science and Engineering, Paavai Engineering College

^{2,3,4}Final Year, Bachelors of Engineering, Computer Science and Engineering

Abstract- Voice-based assessments are increasingly used in online examinations and interview evaluation systems to analyze candidate's communication ability and subject knowledge. Traditional evaluation methods rely on human examiners, which can lead to inconsistencies, bias, and increased evaluation time [2], [3]. Recent research has shown that automatic scoring systems using speech processing, natural language analysis, and multimodal cues can significantly improve assessment efficiency and accuracy [1], [4]. In addition, techniques such as speech emotion recognition help in analyzing vocal characteristics and communication behaviour during spoken responses [5]. This project proposes V-Exam, a voice-based automated evaluation system that uses speech recognition and natural language processing to analyze spoken answers provided by candidates. The system converts voice responses into text, evaluates them for semantic relevance and clarity, and generates an objective performance score. The proposed approach aims to reduce human effort, improve scoring consistency, and enable efficient real-time voice-based assessment.

Key words – Voice-Based Assessment, Speech Recognition, Natural Language Processing, Automated Evaluation, Machine Learning, Real-Time Voice Analysis.

I. INTRODUCTION

Voice-based assessment systems are increasingly being used in online examinations and interview platforms to evaluate candidate's communication skills, knowledge, and speaking ability. Traditional evaluation methods depend on human examiners, which can introduce bias, inconsistency, and increased evaluation time, especially when handling a large number of candidates [2], [3]. With the rapid development of artificial intelligence, automated scoring systems have been developed to analyze spoken responses and provide objective evaluation. Recent research has explored different techniques for automatic scoring of spoken responses using speech processing and machine learning methods. Chen et al. proposed a system for scoring monologue video

interviews by analyzing multimodal cues such as speech and visual features [1]. Similarly, Zechner et al. developed automated scoring techniques for spoken responses in language proficiency assessments using natural language processing approaches [2]. Other studies have investigated automatic evaluation of spontaneous spoken English by analyzing speech features and linguistic characteristics [3].

In addition, deep learning models have been applied to detect off-topic spoken responses and improve the accuracy of automated assessment systems [4]. Speech emotion recognition techniques also play an important role in analyzing vocal expressions and communication behaviour in spoken responses [5]. These technologies help in extracting meaningful patterns from voice data and enable more efficient evaluation of candidate responses.

Motivated by these advancements, this project proposes V-Exam, a voice-based automated evaluation system that uses speech recognition and natural language processing to analyze spoken answers. The system converts voice responses into text, evaluates semantic relevance and clarity, and generates an objective performance score. The proposed approach aims to reduce human effort, improve evaluation consistency, and support real-time voice-based assessment systems

II. RELATED WORKS

Several studies have investigated automated evaluation of spoken responses using speech processing, natural language processing, and machine learning techniques. Early research focused on analyzing spoken responses using linguistic and acoustic features to support automatic scoring systems in language assessment and interview evaluation.

Chen et al. proposed an automated scoring approach for monologue video interviews by combining multimodal cues such as speech characteristics and visual information to evaluate candidate responses [1]. Their work demonstrated that integrating multiple data sources can improve the reliability of automated assessment systems. Similarly, Zechner et al. introduced an automated scoring system for spoken responses in English language proficiency tests using natural language processing and speech recognition techniques, which helped in evaluating fluency and content relevance [2].

Further research has explored the automatic assessment of spontaneous spoken English using speech processing and statistical models. Wang et al. developed a system that analyzes linguistic features, pronunciation quality, and speech patterns to evaluate spoken responses in language learning environments [3]. Their work highlighted the importance of analyzing both linguistic and acoustic features for accurate scoring.

In addition, deep learning methods have been applied to improve automated assessment systems. Wang et al. proposed a convolutional neural network-based approach to detect off-topic spoken responses in automated speaking tests, improving the accuracy of speech evaluation systems [4]. Moreover, speech emotion recognition techniques have been widely studied to analyze emotional cues and vocal characteristics during speech, which can enhance the understanding of speaker behaviour and communication effectiveness [5].

Although these studies demonstrate significant progress in automated speech assessment, many systems still focus on specific aspects such as language proficiency evaluation or multimodal interview analysis. Therefore, there is a need for a voice-based automated evaluation system that can analyze spoken responses, assess semantic relevance, and generate objective performance scores in real time. The proposed V-Exam system aims to address these challenges by integrating speech recognition and natural language processing techniques for efficient voice-based assessment.

III. LITERATURE REVIEW

Automatic evaluation of spoken responses has gained significant attention in recent years due to the increasing use of online interviews and digital assessment platforms. Researchers have explored various techniques such as speech recognition, natural language processing, and machine learning to evaluate spoken responses and improve the efficiency of assessment systems.

Chen et al. developed an automated scoring system for monologue video interviews using multimodal cues such as speech features and visual information [1]. Their study showed that combining different types of features can improve the accuracy and reliability of automated interview evaluation. Zechner et al. proposed an automated scoring approach for spoken responses in English language proficiency assessments using speech recognition and natural language processing techniques [2]. Their work demonstrated that automated systems can effectively evaluate fluency, pronunciation, and content relevance.

Wang et al. investigated automatic assessment of spontaneous spoken English by analyzing speech characteristics and linguistic features [3]. Their research highlighted the importance of evaluating both acoustic and textual features for effective speech assessment. In another study, Wang et al. applied deep convolutional neural networks to detect off-topic spoken responses in automated speaking tests, improving the reliability of scoring systems [4].

Furthermore, Akçay and Oğuz presented a comprehensive review of speech emotion recognition techniques and their applications in speech-based systems [5]. Their research emphasized the role of emotional and vocal characteristics in understanding speaker behaviour during communication.

Although these studies demonstrate the potential of automated speech assessment systems, many existing approaches focus on language proficiency testing or multimodal interview analysis. Therefore, there is a need for a voice-based evaluation framework that focuses on analyzing spoken responses, semantic relevance, and communication clarity. The proposed

V-Exam system addresses this need by integrating speech recognition and natural language processing techniques to provide efficient and objective voice-based assessment.

IV. Existing System

Existing voice-based evaluation approaches typically rely on:

- Manual evaluation of spoken responses by human examiners
- Basic speech-to-text transcription without deeper analysis
- Rule-based keyword matching techniques
- Limited automated language assessment tools

These systems face several limitations such as:

- Time-consuming evaluation process.
- Possibility of human bias and inconsistency in scoring.
- Inability to fully analyze semantic meaning and response quality.
- Lack of intelligent scoring models for spoken responses.
- Limited scalability for handling large numbers of candidates.

Traditional rule-based evaluation methods mainly depend on predefined keywords and patterns, which often fail to capture the true meaning, context, and quality of spoken answers beyond simple matching techniques [3].

Drawbacks

- Manual Evaluation Dependency.
- Lack of Intelligent Semantic Understanding.
- Delayed Response Assessment.
- No Automated Speech Analysis Mechanism.
- Limited Scalability for Large-Scale Examinations.

V. PROPOSED SYSTEM

The proposed system is a software-driven voice-based examination and automated evaluation framework integrating

- Structured question and answer dataset.
- Voice input acquisition through microphone.
- Speech-to-Text conversion using speech recognition models.
- Natural Language Processing for semantic answer analysis.
- Similarity-based response evaluation mechanism.
- Automated scoring and performance analysis.
- Interactive dashboard for result visualization.
- Voice-based feedback and report generation.

Advantages

- Fully automated voice-based evaluation system.
- Real-time speech recognition and response analysis.
- Reduced human bias in scoring and assessment.
- Accurate semantic analysis using NLP techniques.
- Faster evaluation compared to manual assessment.
- Scalable platform for large-scale examinations.
- Improved consistency and reliability in results.

VI. METHODOLOGY

The system methodology consists of the following listed steps:

1. Data Acquisition

Candidate responses are collected through a microphone using a voice-based examination interface. The system records spoken answers as audio input for further processing and evaluation. Voice-based assessment methods are widely used in automated interview and language evaluation systems [1], [3].

2. Speech Processing & Data Preprocessing

The recorded audio is converted into text using speech recognition techniques. The generated transcript is then pre-processed using Natural Language Processing (NLP) methods such as tokenization, stop-word removal, and text

normalization to prepare the data for analysis. Speech recognition and NLP-based preprocessing are commonly used in automated speech evaluation systems [2], [3].

3. Semantic Evaluation

The processed text responses are analyzed to determine their semantic relevance and correctness. The candidate's response is compared with the expected answer using similarity techniques such as cosine similarity or keyword-based matching. These approaches help in identifying whether the response is relevant to the given question and reduce off-topic responses [4].

4. Automated Scoring Mechanism

Based on the similarity score and linguistic features such as clarity, fluency, and completeness, the system generates an automated performance score. Machine learning techniques are used to improve scoring consistency and reduce manual bias in evaluation [1], [2].

5. Result Generation and Reporting

The final evaluation score is displayed through an interactive interface or dashboard. The system can also generate performance reports to help instructors or examiners analyze candidate responses efficiently

6. Speech Behaviour Analysis

In addition to textual evaluation, the system can analyze vocal characteristics such as tone and emotional cues during speech. Speech emotion recognition techniques can enhance understanding of communication behaviour and response confidence in voice-based systems [5].

VII. MODEL DESCRIPTION

A. Speech Recognition Model

The first component of the system is the speech recognition module, which converts spoken responses into text. This module processes the audio signal captured from the microphone and generates a textual transcript of the candidate's answer. Speech recognition models are widely used in automated speech assessment systems to transform spoken language into machine-readable text for further analysis [2], [3].

B. Natural Language Processing Model

After transcription, the generated text is analyzed using Natural Language Processing (NLP) techniques. The system performs tokenization, stop-word removal, and text normalization to prepare the text for semantic analysis. NLP models help identify important keywords and linguistic

C. Semantic Similarity Model

The processed candidate response is compared with the expected answer using semantic similarity techniques such as cosine similarity or keyword-based matching. This model evaluates how closely the candidate's answer matches the correct response in terms of meaning and context. Similarity-based evaluation methods help in detecting off-topic

D. Scoring Model

The final score is generated based on the similarity value and linguistic features extracted from the candidate's response. The scoring mechanism ensures objective and consistent evaluation by reducing human intervention in the assessment process. Automated scoring systems have been widely applied in spoken language evaluation to improve efficiency and reliability [1].

VIII. SYSTEM ARCHITECTURE

The system architecture consists of:

1. Voice Input Module
2. Speech Recognition Module
3. Text Processing Module
4. Semantic Analysis Module
5. Scoring Module
6. Result Display Module

The architecture of the proposed V-Exam system is modular, scalable, and software-centric, enabling efficient voice-based response processing and automated evaluation without requiring specialized hardware infrastructure [3]

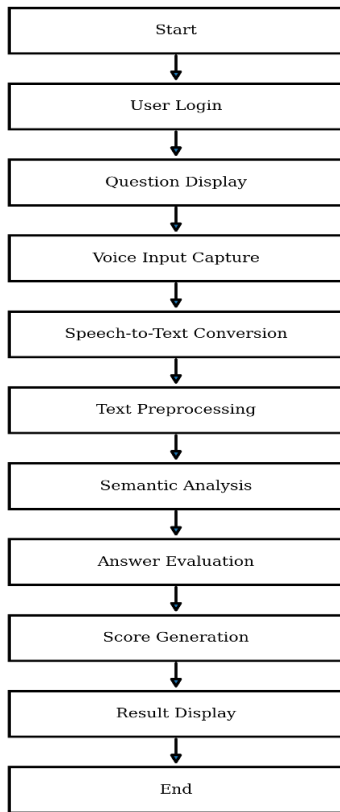


Fig. 1 System Architecture

IX. SYSTEM PROTOCOL

Step	1:	Voice	Capture
Step	2:	Speech	Recognition
Step	3:	Text	Preprocessing
Step	4:	Semantic	Evaluation
Step	5:	Automated	Scoring
Step 6: Result Generation			

X. ALGORITHMS USED

1. Speech-to-Text Conversion Algorithm
 This algorithm converts the user’s spoken input into text format. Tools like Whisper use deep learning-based models to recognize speech patterns and generate accurate textual output. This step is essential for further analysis of voice responses.

2. Text Preprocessing Algorithm

This algorithm cleans and prepares the text for analysis. It includes steps such as tokenization, stop-word removal, stemming, and normalization. These techniques ensure that the text is structured and meaningful for processing.

3. Natural Language Processing (NLP) Algorithm
 NLP algorithms are used to understand the meaning and context of the user’s response. Techniques such as part-of-speech tagging and semantic analysis help in evaluating the relevance of the answer.

4. Cosine Similarity Algorithm
 This is the main scoring algorithm used in your project. It compares the user’s answer with the reference answer by calculating similarity between text vectors. The closer the similarity score to 1, the more accurate the answer is.

Formula (optional if needed):

$$\text{Similarity} = \frac{A \cdot B}{(\|A\| \times \|B\|)}$$

5. TF-IDF (Term Frequency – Inverse Document Frequency)
 This algorithm converts text into numerical vectors based on word importance. It gives higher weight to important words and lower weight to common words, improving similarity calculation accuracy.

XI. RESULTS AND DISCUSSION

The proposed V-Exam system was developed to evaluate spoken responses automatically using speech recognition and natural language processing techniques. The system processes candidate voice input,

During testing, the system successfully captured voice responses through the microphone and converted them into text using the speech recognition module. The text processing module effectively cleaned and prepared the generated transcript for semantic analysis.

The results demonstrate that the system can evaluate spoken responses efficiently and provide consistent scoring without manual intervention. The automated scoring mechanism reduces evaluation time and improves fairness by minimizing human bias. Additionally,

Overall, the experimental results indicate that the V-Exam system provides reliable voice-based evaluation and improves the efficiency of assessment processes. The integration of speech recognition and semantic analysis enables accurate response evaluation and supports scalable voice-based examination platforms.

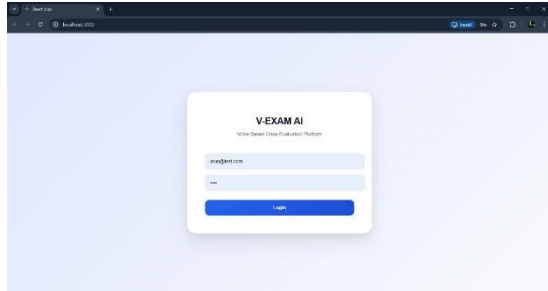


Fig. 2 Login Page

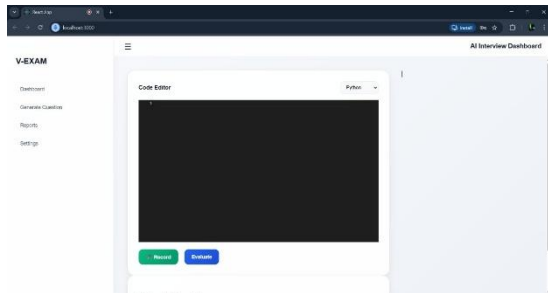


Fig. 3 Coding and Explanation Page

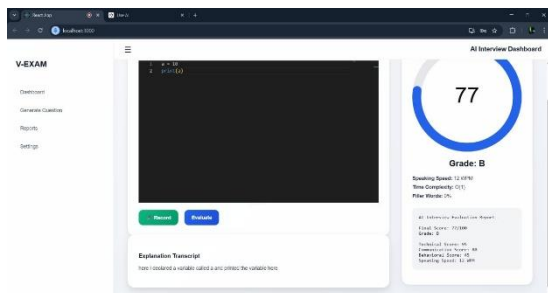


Fig. 4 Generated Report

Feature	Importance Value
Semantic Relevance	0.38
Speech Clarity	0.34
Response Completeness	0.28

Fig. 5 Response Evaluation Factors

XII. CONCLUSION

This project presented V-Exam, A voice-based automated evaluation system designed to assess candidate responses using speech recognition and natural language processing techniques. The system captures spoken answers, converts them into text, and analyzes the responses to determine their semantic relevance, clarity, and completeness. By integrating automated speech processing and intelligent scoring mechanisms, the system provides objective and consistent evaluation of voice-based responses.

The proposed approach reduces the dependency on manual assessment and minimizes human bias in the evaluation process. It also improves efficiency by enabling faster and scalable assessment of multiple candidates in online examination and interview environments. Overall, the V-Exam system demonstrates the potential of voice-based technologies in creating intelligent and reliable automated evaluation platforms for modern digital assessment systems.

REFERENCES

- [1] Automatic Scoring of Monologue Video Interviews Using Multimodal Cues. Chen, L., Zhao, R., Wang, L., and Pietikäinen, M. In Proceedings of Inter-speech, 2016, pp. 2592–2596.
- [2] Automatic Scoring of Speaking Items in an Assessment for Teachers of English as a Foreign Language. Zechner, K., Higgins, D., Xi, X., and Williamson, D. In Proceedings of the NAACL Workshop on Innovative Use of NLP for Building Educational Applications, 2014, pp. 45–56.
- [3] Towards Automatic Assessment of Spontaneous Spoken English. Wang, Y., Gales, M. J. F., and Knill, K. Speech Communication, vol. 98, 2018, pp. 49–63.
- [4] Automatic Detection of Off-Topic Spoken Responses Using Very Deep Convolutional Neural Networks. Wang, X., Liu, J., and Xu, B. In Proceedings of Inter-speech, 2019, pp. 2723–2727.

- [5] Speech Emotion Recognition: A Review of Methods and Applications. Akçay, M. B., and Oğuz, K. *IEEE Access*, vol. 8, 2020, pp. 19147–19164.
- [6] A. Metallinou and S. Narayanan, “Automatic Speech Assessment Using Deep Neural Networks,” *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 892–903, 2022.
- [7] J. Li, H. Xu, and T. Schultz, “Deep Learning Approaches for Automated Spoken Language Assessment,” *IEEE Access*, vol. 10, pp. 45876–45889, 2022.
- [8] S. D’Mello and R. Graesser, m“Multimodal Learning Analytics for Speech-Based Assessment Systems,” *IEEE Transactions on Learning Technologies*, vol. 15, no. 3, pp. 356–367, 2022.
- [9] M. Qian, Y. Wang, and H. Meng, “Automatic Evaluation of Spoken Responses Using Neural Network Models,” *Computer Speech & Language*, vol. 73, 2023. (Springer/Elsevier indexed)
- [10] Kato and Y. Yamamoto, “Automatic Scoring of Spoken Language Proficiency Using Deep Learning,” *Speech Communication*, vol. 148, pp. 45–57, 2023.
- [11] L. Zhang, X. Li, and Y. Chen, “Speech-Based Assessment Systems Using Transformer Models,” *IEEE Access*, vol. 11, pp. 103245–103256, 2023.
- [12] H. Kim and J. Park, “Speech Emotion Recognition for Intelligent Voice-Based Evaluation Systems,” *IEEE Transactions on Consumer Electronics*, vol. 69, no. 2, pp. 231–240, 2023.
- [13] R. Patel and K. Shah, “Automatic Assessment of Spoken Responses Using NLP and Machine Learning,” *Neural Computing and Applications*, Springer, vol. 36, pp. 2145–2159, 2024.
- [14] Y. Liu, H. Wang, and P. Fung, “Transformer-Based Models for Automated Speech Assessment,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 1502–1514, 2024.
- [15] M. Rodriguez and S. Gupta, “AI-Based Voice Evaluation Systems for Online Interview Platforms,” *IEEE Access*, vol. 13, pp. 21478–21491, 2025.