CONVERS AI – CLEVER MIX OF COMMUNICATION WITH AI

Madhura Pande[1], Munja Sakhare[2], Payal Shirbhate[3], Sharayu Dubey[4], Mrinmay Samarth[5], Dr. S.S Asole[6]
1,2,3,4,5 are students of Department Computer Science and Engineering and 6 Faculty of Babasaheb Naik College of Engineering, Pusad-445204, Maharashtra, India

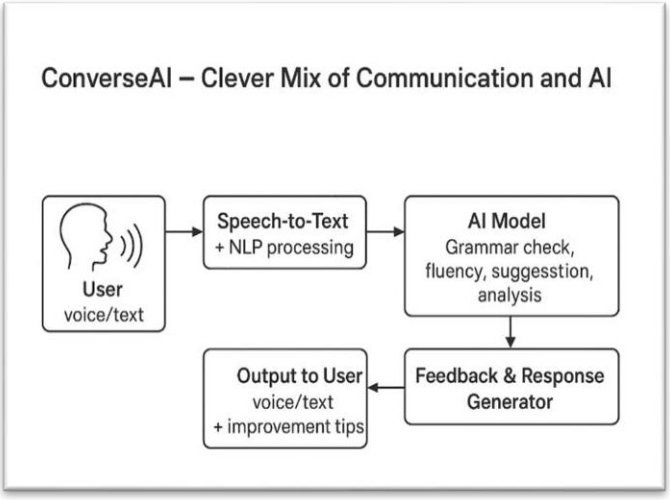
Abstract :
The increasing demand for effective communication skills in both academic and professional domains has led to the development of intelligent tools for language proficiency evaluation. Traditional communication assessments rely heavily on human evaluators, which are often time-consuming, subjective, and resource-intensive. Existing digital platforms provide practice-based support but lack structured evaluation and personalized feedback. This paper presents the design and implementation of ConversAI, a voice-based AI chatbot that provides an automated and objective evaluation of communication skills. Unlike traditional systems, the proposed framework integrates automated speech recognition (ASR), natural language processing (NLP), sentiment analysis, and scoring algorithms to assess both linguistic aspects (grammar, vocabulary, coherence) and paralinguistic features (fluency, pronunciation, confidence, and tone). The system generates instant scores, detailed reports, and constructive feedback, enabling continuous improvement. By offering automation, scalability, and unbiased evaluation, ConversAI serves as a cost effective and reliable solution for educational institutions, corporate organizations, and recruitment processes. It bridges the gap between practice tools and formal assessment methods, making communication skill evaluation smarter, faster, and more accessible.

***Keywords**: Automatic Speech Recognition (ASR) , Fluency and pronunciation evaluation , Automated feedback , Interview preparation*

1. Introduction

Effective communication skills are essential in both academic and professional domains, influencing career success,academic performance, and interpersonal relationships. Traditional methods of evaluating communication, such as interviews or language proficiency tests, often rely on human evaluators. While effective, these methods are time-consuming, subjective, and resource- intensive.

With the rise of artificial intelligence (AI) and natural language technologies, automated systems now offer opportunities to make communication assessment smarter, faster, and more objective. Existing digital platforms primarily provide practice-based support but lack structured evaluation and detailed, personalized feedback.



To address this gap, we propose **ConverseAI**, a voice-based AI chatbot designed to evaluate communication skills automatically. The system integrates Automatic Speech Recognition (ASR), Natural Language Processing (NLP), sentiment analysis, and scoring algorithms to assess linguistic features (grammar, vocabulary, coherence) as well as paralinguistic features (fluency, pronunciation, confidence, tone). Unlike traditional assessments, ConverseAI delivers instant scores, comprehensive reports, and actionable feedback, enabling learners to continuously improve.

By offering automation, scalability, and unbiased evaluation, ConverseAI can serve as a cost-effective solution for educational institutions, corporate organizations, and recruitment processes. This project aims to bridge the gap between practice tools and formal communication assessments, making skill evaluation more accessible and reliable.

2. Literature Review

2.1 Automatic Speech Recognition (ASR) in Language Learning

ASR forms the backbone of automated speaking assessment tools, enabling transcription of learner speech for further analysis. Modern end-to-end ASR systems achieve high recognition accuracy but still face challenges with spontaneous, accented, and noisy speech. Studies show that ASR-based feedback provides moderate improvements in pronunciation when combined with explicit corrective strategies (Hu et al., 2023). Despite its limitations, ASR remains a reliable foundation for interview-practice chatbots.

2.2 Automated Pronunciation and Fluency Assessment

Research in automated pronunciation scoring has shown promising results in identifying phoneme-level errors through acoustic modeling and alignment techniques (Zhang et al., 2023). While effective at detecting segmental errors, suprasegmental aspects such as stress and intonation remain challenging. Similarly, fluency assessment relies on speech rate, pause distribution, and continuity measures extracted from ASR outputs. Targeted fluency feedback has been shown to improve learner confidence, though distinguishing between meaningful pauses and disfluency remains difficult.

2.3 Grammar and Content Feedback

Grammar error detection typically relies on ASR transcripts followed by NLP-based correction modules. While effective for typed text, grammar checkers are sensitive to transcription errors. Recent studies integrate ASR confidence scores or lattice- based correction to reduce false positives (Huang & Li, 2022). Additionally, large language models now enable semantic evaluation, providing insights into content relevance and appropriateness in interview responses (Lee et al., 2024).

2.4 Conversational Agents for Language Learning

Conversational agents have been widely applied in education, offering learners role- play practice and interactive feedback. Reviews highlight that such agents enhance motivation, speaking confidence, and engagement (Chou et al., 2023). For interview training, prototypes such as MACH and AI-powered dialogue systems simulate realistic interview scenarios and provide multimodal feedback (Nguyen et al., 2024). However, the pedagogical design of feedback remains crucial, as overly detailed corrections can overwhelm learners.

2.5 Effectiveness and Challenges

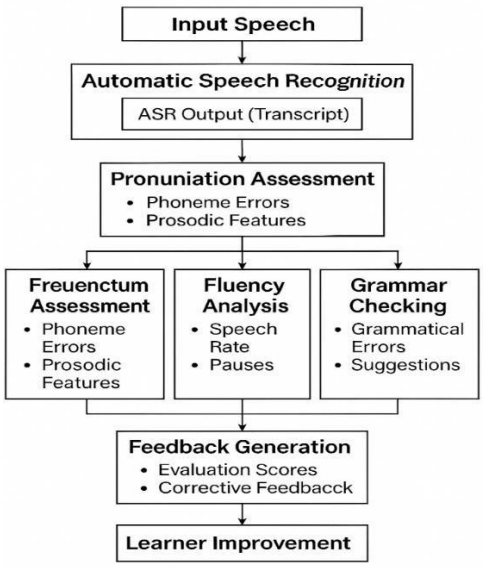
Meta-analyses confirm that ASR-based tools produce moderate improvements in pronunciation and oral performance, while also reducing language anxiety (Hu et al., 2023). Challenges persist in three main areas: cascading ASR errors affecting grammar/content feedback, fairness issues in pronunciation scoring across different L1 backgrounds, and limited evidence on long- term transfer to real-world interview success. Privacy concerns related to storing and processing voice data are also significant.

2.6 Research Gaps and Opportunities

Existing systems mainly focus on controlled speech rather than spontaneous interview-style responses. There is also a lack of context-aware pause interpretation, which could better distinguish thoughtful silence from disfluency. Furthermore, grammar correction systems need to integrate ASR uncertainty, and scoring models should account for L1 variation to improve fairness. Addressing these gaps positions ConverseAI as a novel contribution that integrates multi- dimensional feedback for real-time interview practice.

This framework illustrates the typical pipeline used in AI-driven spoken language assessment systems. The process begins with input speech, which is converted into text using Automatic Speech Recognition (ASR). The ASR transcript is then analyzed through multiple components:

1. Pronunciation Assessment – Detects phoneme-level errors and prosodic features (intonation, stress, rhythm).
2. Fluency Analysis – Evaluates speech rate and pauses to measure natural flow.
3. Grammar Checking – Identifies grammatical errors and provides suggestions.
4. **Frequentum(Frequency/Pronunc iation) Assessment** – Examines repeated **phoneme errors** and prosodic patterns.
5. **Feedback Generation** – Produces



evaluation scores and gives corrective feedback to the learner.

Finally, this cycle contributes to learner improvement, aligning with studies in computer-assisted language learning (CALL) and AI-based communication training.

3. Methodology

The methodology of the proposed system involves several phases, starting from user speech input to learner improvement through AI-driven evaluation and feedback. The complete workflow is shown in Figure**.**

Steps of Methodology

1. **Input Speech Collection**
   - The learner provides spoken input (e.g., during interview practice or communication training).
   - The system records this audio for further processing.
2. **Automatic Speech Recognition (ASR)**
   - Converts speech into text using ASR engines.
   - Outputs a transcript that is used for further analysis.
3. **Feature Extraction & Analysis**
   - Pronunciation Analysis: Detects phoneme-level errors and evaluates prosodic features such as stress and intonation.
   - Fluency Analysis: Measures speech rate, pauses, and hesitations.
   - Grammar Analysis: Identifies grammatical mistakes and suggests corrections.
4. **Evaluation & Scoring**
   - The system generates quantitative scores (accuracy, fluency, grammar, etc.).
   - Uses AI/ML models to benchmark the learner's performance.
5. **Feedback Generation**
   - Provides corrective feedback in terms of suggestions for better pronunciation, grammar corrections, and fluency improvement.
   - Learners receive both scores and qualitative feedback.
6. **Learner Improvement**
   - Continuous usage of the system helps learners improve their spoken communication skills through iterative practice and feedback.

4. System Architecture

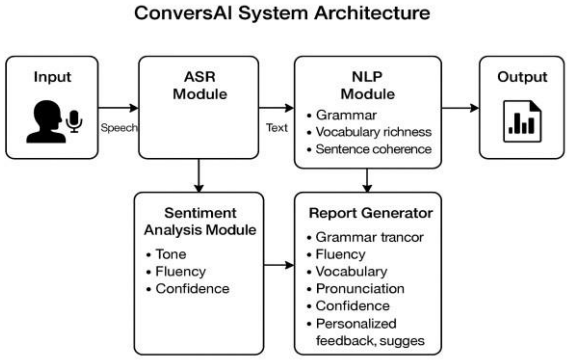


Fig.1:System Architecture

The proposed *ConversAI* architecture consists of five key components that work together to assess communication skills (Fig. 1).

Input Layer:
The system begins with the student providing spoken responses through a microphone. The audio input is processed as the raw data for assessment.

Automatic Speech Recognition (ASR):
The speech is transcribed into text using ASR technologies such as Google Speech-to-Text or Mozilla DeepSpeech. ASR serves as the foundation of the system, since accurate transcription is essential for subsequent analysis. Previous research has shown that modern ASR systems achieve near- human accuracy under controlled conditions [2].

Natural Language Processing (NLP) Module:
The transcribed text undergoes grammatical, syntactic, and semantic analysis. This includes evaluation of
grammar correctness, vocabulary richness, and sentence coherence, which are central to communication skills [7].

Sentiment Analysis Module: To assess communication quality beyond textual accuracy, sentiment analysis is employed. This module evaluates tone, fluency, and confidence, based on acoustic prosody and linguistic cues [4]. This step ensures that both **content** and **delivery** are measured.

Report Generator:
The results from NLP and sentiment analysis are compiled into a structured report. The report provides detailed scores for grammar, fluency, vocabulary, pronunciation, and confidence. Personalized feedback and improvement suggestions are also generated to guide students in self- learning.

Output Layer:
Finally, the generated report is delivered to the student in textual and graphical form.
This report is designed to be easy to interpret and helps students identify their strengths and areas for improvement.

5. Conclusion

The proposed AI-based system provides an efficient, scalable, and objective solution for enhancing spoken communication skills. By integrating speech recognition, language analysis, and personalized feedback, it ensures continuous improvement and adaptability. With further refinement, it can act as an intelligent tutor for language learning, interview preparation, and communication training.

6. References

[1] World Economic Forum, Future of Jobs Report, WEF, 2020.
[2] Hosseini, R., et al., "Artificial intelligence in communication skill training," Journal of Educational Technology, 2018.
[3] Huang, X., Baker, J., & Reddy, R., "A historical perspective of speech recognition," Communications of the ACM, vol. 57, no. 1, pp. 94–103, 2019.
[4] Cambria, E., & White, B., "Jumping NLP curves: A review of natural language processing research," IEEE Computational Intelligence Magazine, vol. 9, no. 2, pp. 48–57, 2014.
[5] HireVue, "AI-powered hiring and interview assessments," [Online]. Available: https://www.hirevue.com
, 2020.
[6] Google, "Interview Warmup," [Online].
Available:
https://grow.google/interview-warmup
, 2022.
[7] Jurafsky, D., & Martin, J. H., Speech and Language Processing, 3rd ed., Pearson, 2021.
[8] Winkler, R., & Söllner, M., "Unleashing the potential of chatbots in education: A state-of-the-art analysis," Academy of Management Proceedings, 2018(1), 15903.