ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug1065

Speech-to-Text AI for Improving English Pronunciation in ESL Learners

Dr. T. Prakash[1*]; S. Kausalya[2]

[1]Librarian; [2]Assistant Librarian

[1,2]Department of Library, Nandha College of Technology, Erode, Tamil Nadu – 638052, India

Corresponding Author: Dr. T. Prakash[*]

Publication Date: 2025/08/27

Abstract: English as Second Language (ESL) learners often struggle with pronunciation, which can hinder academic success and social integration. This study investigates the effectiveness of a speech-to-text artificial intelligence (AI) system in improving pronunciation accuracy among ESL learners. Using a phoneme-matching approach, the system provided real-time corrective feedback to students in semi-urban learning environments. Data were collected through pre- and post-tests measuring accuracy, precision, recall, and F1-score. Results revealed a 15% improvement in pronunciation accuracy, supported by consistent gains across all performance metrics. Learners also demonstrated increased confidence and sustained engagement, highlighting the motivational value of instant AI-based feedback. These findings suggest that speech-to-text AI can complement traditional instruction by offering personalized and continuous pronunciation training. Future research should explore long-term retention and integration with immersive technologies such as virtual and augmented reality.

Keywords: *Speech-to-Text, ESL Pronunciation, Artificial Intelligence, Natural Language Processing (NLP), Phoneme Feedback.*

How to Cite: Dr. T. Prakash; S. Kausalya (2025), Speech-to-Text AI for Improving English Pronunciation in ESL Learners. *International Journal of Innovative Science and Research Technology*, 10(8), 1341-1343. https://doi.org/10.38124/ijisrt/25aug1065

I. INTRODUCTION

ESL learners often struggle with pronunciation due to limited exposure to native speakers and a lack of personalized feedback, which can hinder their academic and social integration. AI offers a solution by providing customized, immediate, and targeted feedback on pronunciation. Such tools have the potential to transform language learning by supporting learners with tailored, interactive guidance.

II. LITERATURE REVIEW

Research in language education emphasizes that targeted pronunciation feedback and phoneme-level analysis play a critical role in improving learner outcomes. Studies highlight that interactive approaches—such as video-based communication tasks—support better rhythm, articulation, and intonation in spoken English (Brown, 2019; Li, 2020). However, much of this work has been conducted in urban contexts, leaving semi-urban and rural learners underrepresented. In these regions, limited exposure to native speakers and authentic communication environments continues to hinder effective pronunciation development (Rahman & Devi, 2022). Another limitation is the scarcity of longitudinal studies, making it difficult to determine whether improvements in pronunciation are sustained over time (Lee, 2018).

- *AI in Language Learning*

Artificial intelligence (AI) has emerged as a valuable tool in language learning by offering adaptive feedback, customized learning experiences, and real-time progress tracking. Neural network–based systems can identify subtle errors in articulation and provide immediate corrective input, enabling learners to refine pronunciation more efficiently than with delayed classroom feedback (Smith & Johnson, 2020). AI also facilitates independent learning, allowing students to practice at their own pace while still benefiting from individualized guidance (Kumar, 2021). By extending opportunities for practice beyond traditional instruction, AI complements teachers' roles and enhances learner autonomy.

- *Gaps in Existing Research*

Despite progress in AI-assisted pronunciation training, significant challenges remain. Research in semi-urban and rural contexts is still limited, where learners often face barriers such as reduced access to fluent English speakers (Rahman & Devi, 2022). Furthermore, immersive technologies such as Virtual Reality (VR) and Augmented Reality (AR) have not yet been fully integrated into language

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug1065

learning research (Kumar, 2021). The lack of longitudinal investigations further restricts understanding of whether AI-supported improvements are retained in the long run (Thomas & Lee, 2020). Addressing these gaps will require a combination of AI-driven feedback, cognitive learning theory, and effective pedagogical strategies to optimize pronunciation outcomes.

III. METHODOLOGY

➢ *Speeches-to-Text Framework*

A customized automatic speech recognition (ASR) model was designed with deep learning methods. Convolution Neural Networks (CNN) was utilized for extracting sound features, while Recurrent Neural Networks (RNN) managed sequential data. The model converted speech inputs into Mel-Frequency Cepstral Coefficients (MFCCs) and mapped them to their linguistic counterparts, preserving both context and natural intonation.

➢ *Phoneme Evaluation Process*

Following transcription, the system analyzed phonemes against a reference dataset of standard pronunciations. Errors were grouped as substitutions, omissions, or insertions. Learners received corrective input through text prompts, visual feedback, and synthetic voice examples. This multi-sensory approach encouraged continuous reinforcement and better retention of correct speech habits.

➢ *Participants and Data*

The study included 60 students from semi-urban institutions. Over a four-week period, learners submitted recordings of sentences, words, and spontaneous speech. Performance was assessed before and after training to measure accuracy, fluency, and self-confidence.

➢ *Classroom Application*

The AI system was incorporated into daily lessons. Students engaged with guided pronunciation drills, self-corrective tasks, and interactive exercises. Teachers tracked learners' progress through an analytics dashboard, allowing targeted guidance and personalized study plans.

IV. RESULTS & DISCUSSION

A. Performance Metrics

Table 1 Performance Metrics

Metric	Before AI	After AI
Accuracy	75%	90%
Precision	70%	88%
Recall	80%	92%
F1-Score	75%	90%

➢ *Interpretation:*

The comparison clearly demonstrates that integrating AI-based speech recognition systems produced measurable improvements in learners' pronunciation performance.

- *Accuracy*

✓ Before AI: 75%
✓ After AI: 90%

This 15% increase shows that students articulated words more correctly after receiving AI-supported guidance. The system's instant feedback allowed learners to quickly identify errors and adjust their speech.

- *Precision*

✓ Before AI: 70%
✓ After AI: 88%

Precision reflects how consistently correct the recognized words were. The 18% gain indicates that students not only pronounced words accurately but also reduced inconsistency in articulation.

- *Recall*

✓ Before AI: 80%
✓ After AI: 92%

Recall represents the ability of learners to reproduce correct sounds when prompted. The 12% improvement demonstrates that learners were able to retain corrective feedback and apply it more effectively in practice.

- *F1-Score*

✓ Before AI: 75%
✓ After AI: 90%

The balanced rise in F1-score shows simultaneous improvement in both precision and recall. Learners not only made fewer errors but also applied corrections consistently across different speech tasks.

B. Graphical Insights

The improvements illustrated in the performance graph confirm that speech-to-text AI significantly enhances ESL learners' pronunciation skills. Gains across all four key metrics—Accuracy, Precision, Recall, and F1-score—highlight the system's effectiveness as a learning companion.

Immediate phoneme-level feedback emerged as a critical factor: learners were able to detect and correct errors in real time, which shortened the learning cycle compared with delayed classroom feedback. This process allowed them to internalize correct pronunciation patterns faster and with greater confidence.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug1065

In addition to measurable results, students reported feeling more engaged and motivated. The sense of achievement provided by immediate recognition encouraged continuous practice and fostered a positive learning environment.

V. ANALYSIS AND EXPERIMENTS

➢ *Phoneme-Level Feedback*

one of the most valuable contributions of the AI system was its ability to provide immediate feedback at the phoneme level. Unlike classroom corrections, which are often delayed or generalized, the AI tool delivered precise, real-time guidance. This shortened the learning process and allowed learners to quickly adopt correct pronunciation patterns.

➢ *Learner Engagement and Motivation*

The availability of instant recognition and feedback encouraged students to participate more actively. Receiving confirmation for correct pronunciation boosted their confidence and reduced learning anxiety. This reinforcement cycle here progress motivated additional practice—helped learners steadily improve their performance.

➢ *Pedagogical Implications*

The study suggests that AI applications can function as personalized pronunciation coaches. Rather than replacing the teacher, these systems provide complementary support by bridging the gap between classroom instruction and independent practice. As a result, learning becomes more individualized, efficient, and accessible to a wider range of learners.

➢ *Role of AI in ESL Learning*

The findings validate the assumption that speech-to-text AI enhances pronunciation accuracy, precision, recall, and F1-score. More importantly, the technology transforms pronunciation training into a learner-centered, interactive, and data-driven process. Compared with traditional approaches, this method accelerates skill development and ensures better learner outcomes.

VI. CONCLUSION & FUTURE SCOPE

➢ *Conclusion:*

This study demonstrates that speech-to-text AI systems significantly enhance English pronunciation among ESL learners by providing instant, phoneme-level feedback. Learners not only reduced pronunciation errors but also gained confidence and sustained motivation through continuous interaction with the system. Importantly, AI does not replace teachers but reinforces classroom instruction by enabling personalized, independent practice.

➢ *Future Scope:*

Future studies should examine long-term retention of pronunciation improvements and extend research to rural and semi-urban contexts where access to expert instruction is limited. Integrating AI with immersive technologies such as Virtual Reality (VR) and Augmented Reality (AR) may further enhance learner engagement. Expanding sample sizes and conducting longitudinal studies will provide deeper insights into the sustained impact of AI-assisted pronunciation training.

REFERENCES

[1] Brown, J., & Miller, S. (2019). Video-based communication tasks in ESL pronunciation. *International Journal of Education, 10*(2), 23–35.
[2] Garcia, L., & Patel, R. (2021). Adaptive AI systems for ESL learning. *Journal of Applied Linguistics and AI, 8*(1), 50–67.
[3] Kumar, P. (2021). Virtual reality in second language education. *Language Technology Review, 5*(3), 77–90.
[4] Lee, A. (2018). Long-term effects of pronunciation training in ESL learners. *TESOL Quarterly, 52*(4), 1020–1035.
[5] Rahman, F., & Devi, K. (2022). Barriers to ESL learning in semi-urban contexts. *Asian Journal of English Studies, 14*(2), 88–101.
[6] Smith, M., & Johnson, R. (2020). AI for ESL pronunciation. *Journal of Language Research, 15*(2), 50–65.
[7] Thomas, J., & Lee, S. (2020). Retention of pronunciation gains in AI-assisted learning. *Educational Technology & Society, 23*(4), 115–128.
[8] Wang, H., & Chen, Y. (2020). Neural networks for speech recognition in ESL learners. *Computers & Education, 149*, 103809.