

Enhancing Pronunciation Skills through AI-Based Speech Recognition Technologies

Rousan Ara¹

Research Assistant, Bangladesh Center for Research and Planning (BCRP)
Email: rumi.bd10@gmail.com

Mohammad Shah Alam Chowdhury²

Professor, Department of English, Dhaka International University
Email: shahalam.eng@diu.ac
Orcid ID: 0000-0003-0348-5739

Mohammad Mehedi Hasan³

Assistant Professor, Department of Political Science, Dhaka International University
Email: mehedihasan.polit@diu.ac
Orcid ID: 0000-0002-6357-7742

Keywords:

*Artificial Intelligence,
Speech Recognition
Technology,
Pronunciation
Learning, EFL,
Language Education,
Computer-Assisted
Language Learning*

Abstract: Pronunciation is one of the most difficult areas of second language acquisition and more especially for EFL learners. Typical classroom settings can be difficult to provide individual pronunciation feedback and corrective feedback. The recent developments of Artificial Intelligence (AI) especially language learning through AI speech recognition have revolutionized the way language is learned, providing instant, personalized and data-driven pronunciation feedback. This paper proposes to examine how well AI speech recognition technology can improve the pronunciation skills of learners. The study is based on previous research, theories on second language acquisition, and current advances in education technology to explore the role of AI applications in facilitating pronunciation accuracy, learner autonomy, motivation, and confidence. The paper proposes that in the field of pronunciation learning, continuously assessing and providing immediate corrective feedback on students' pronunciation through AI-based speech recognition technology can generate adaptive learning experiences and effectively enhance pronunciation learning outcomes. However, the potential for educating with them is only realized if a number of issues are tackled, including technology, accent bias, privacy and accessibility. The results indicate that the use of AI pronunciation tools in language teaching could enhance the learning atmosphere and be effective when combined with a classical language teaching method.

Introduction

Artificial Intelligence (AI) has transformed various industries such as health care, finance, transportation, and education. In the field of language education, AI technologies have introduced new and innovative approaches to teaching and learning, especially in areas that require individualized practice and personalized feedback (Zawacki-Richter et al., 2019). The century of the twenty-first is the age of

the period in which Artificial Intelligence (AI) has become one of the most prominent technological innovations (Ara et al., 2026). AI's potential for language learning is particularly significant in the area of speech recognition technology, as it enables learners to receive immediate feedback on their speaking performance and pronunciation development (Godwin-Jones, 2018). Pronunciation is considered an essential component of communicative competence because accurate and intelligible pronunciation contributes to effective communication, listening comprehension, and successful interaction. According to Derwing and Munro (2015), pronunciation instruction should focus on improving intelligibility rather than achieving native-like speech. Although pronunciation plays an important role in foreign language acquisition, it has often received limited attention in traditional language classrooms (Levis, 2018). Due to various classroom challenges, including limited instructional time, large class sizes, and diverse learner needs, providing individualized pronunciation training is often difficult for teachers. AI-based speech recognition systems offer possible solutions to these challenges by creating opportunities for personalized learning and continuous pronunciation practice (Warschauer & Healey, 1998). From automated speech evaluation systems to language learning applications, these technologies use advanced algorithms to analyze speech patterns, identify pronunciation errors, and provide targeted corrective feedback (Neri et al., 2008). Such technologies allow learners to practice independently beyond the classroom environment and support the development of learner autonomy. Recent studies have shown growing academic interest in the application of AI tools for language learning, particularly in improving pronunciation accuracy, learner engagement, and overall language proficiency (Dizon, 2020). Although research findings demonstrate promising outcomes, concerns remain regarding speech recognition accuracy, algorithmic bias, recognition of different accents, and ethical issues related to learner data privacy (Zawacki-Richter et al., 2019).

Research Objectives

1. To explore the use of pronunciation learning technologies based on Artificial Intelligence (AI) speech recognition.
2. To explore how effectively the pronunciation feedback from AI can enhance accuracy.
3. To investigate student attitudes to the use of AI for pronunciation training.
4. To come up with problems related to the use of speech recognition technology in language learning.

Research Questions

1. How does speech recognition technology using Artificial Intelligence help with pronunciation?
2. What effect does instant feedback have on the accuracy of learners' pronunciation?
3. What are learners' perceptions of the use of AI for pronunciation learning?
4. What are their limitations and challenges in terms of effectiveness when using speech recognition technologies?

Literature Review

Pronunciation is an important element in language acquisition because it affects language comprehension and the effectiveness of communication. Derwing and Munro (2015) argue that intelligible pronunciation is more significant than achieving native-like pronunciation. Even learners with strong grammar and vocabulary knowledge can experience communication difficulties if their pronunciation ability is limited (Derwing & Munro, 2015). The making use of intelligent tutoring systems, automated writing assessment tools, speech recognition systems, machine translation software, and generative AI models can facilitate a more customized learning experience and offer real-time feedback (Ara et al., 2026). Pronunciation teaching has developed from traditional repetition-based approaches toward communicative approaches that emphasize intelligibility. However, pronunciation has often been neglected in language classrooms, and teachers may not always have sufficient time or resources to provide individualized phonetic feedback (Levis, 2018). Previous research has shown that targeted pronunciation instruction can significantly improve learners' speaking proficiency. Continuous practice, corrective feedback, and learner self-monitoring are essential components of successful pronunciation development (Pennington & Rogerson - Revell, 2019). These requirements are closely connected with the functions of AI-based speech recognition technologies, which can provide repeated practice and individualized feedback. Speech recognition technology refers to the ability of computer systems to recognize, analyze, and process human speech. Early speech recognition systems faced several challenges because of computational limitations and lower accuracy rates. However, advances in machine learning, deep learning, and natural language processing have greatly improved the effectiveness of modern speech recognition systems (Jurafsky & Martin, 2023). Modern speech recognition systems use neural networks trained on large speech datasets to analyze pronunciation patterns, recognize spoken language, and provide detailed feedback. These technologies are increasingly being adopted in educational applications to support language learning and pronunciation development (Liakin et al., 2015). AI-

driven speech recognition has introduced adaptive learning by adjusting activities according to individual learner performance. These systems can identify specific pronunciation problems and recommend personalized exercises, creating more effective and learner-centred instruction (Godwin-Jones, 2018). The field of Computer-Assisted Language Learning (CALL) has changed significantly over the past decades. Early CALL systems mainly focused on grammar exercises and vocabulary practice, while modern AI-supported CALL environments provide interactive, adaptive, and personalized learning experiences (Warschauer & Healey, 1998). Researchers have identified several advantages of AI-supported language learning technologies, including immediate feedback, learner autonomy, increased practice opportunities, and individualized instruction. AI systems can monitor learners' progress over time and provide customized learning pathways (Chapelle, 2001). One of the most important contributions of AI-based speech recognition technologies in CALL is their ability to provide personalized pronunciation feedback on a large scale, allowing learners to improve their speaking skills independently (Neri et al., 2008). Many studies have investigated the effectiveness of AI-based speech recognition technologies in pronunciation learning. Research findings generally indicate improvements in pronunciation accuracy, fluency, and learner confidence through the use of computer-assisted pronunciation training tools (Neri et al., 2008). Students using AI-supported pronunciation tools often demonstrate improvement in segmental features, such as vowel and consonant pronunciation, as well as suprasegmental features, including stress, rhythm, and intonation (Levis, 2018). A major benefit of speech recognition technology is immediate feedback. AI systems provide instant evaluation, allowing learners to identify and correct errors more quickly than in traditional learning environments (Liakin et al., 2015). AI pronunciation practice can also reduce learners' speaking anxiety because students are able to practice privately and repeatedly without fear of negative evaluation from teachers or classmates (Godwin-Jones, 2018). However, some limitations remain. Speech recognition systems may experience difficulty processing different accents, background noise, and non-standard pronunciation patterns, which may sometimes lead to inaccurate feedback (Derwing & Munro, 2015).

Theoretical Framework

Krashen's Input Hypothesis: According to Krashen's Input Hypothesis, comprehensible input is one of the most important factors in successful language acquisition. The theory suggests that learners acquire language when they receive input slightly above their current level of competence (Krashen, 1985). AI-based speech recognition technologies support this theory by providing learners with

authentic pronunciation models, repeated exposure, and corrective feedback. Through continuous interaction with AI systems, learners receive meaningful input and opportunities for pronunciation improvement. Schmidt's Noticing Hypothesis Schmidt (1990) explains that conscious awareness of linguistic features is essential for language learning. Learners must notice differences between their own language production and the target language form in order to develop accuracy. Speech recognition technologies support the noticing process by identifying pronunciation errors and highlighting differences between learner pronunciation and target pronunciation. Visual feedback, audio comparison, and correction suggestions increase learner awareness and encourage self-correction. Vygotsky's Sociocultural Theory: Vygotsky's Sociocultural Theory emphasizes the importance of interaction, mediation, and scaffolding in the learning process. According to Vygotsky (1978), learners develop through support provided within the Zone of Proximal Development (ZPD). AI-powered pronunciation tools can function as technological scaffolding by offering adaptive feedback and personalized activities. These tools allow learners to practice at their own level while receiving appropriate guidance.

Self-Regulated Learning Theory

Self-regulated learning emphasizes learners' ability to control, monitor, and evaluate their own learning processes. Zimmerman (2002) explains that successful learners actively set goals, monitor their progress, and adjust learning strategies. AI-based pronunciation software promotes self-regulated learning by providing performance analysis, progress tracking, and customized learning paths. These features encourage learner autonomy and help students take responsibility for their language development.

Research Methodology

The present study adopts a qualitative research design based on a literature-based approach. It synthesizes findings from peer-reviewed journal articles, conference papers, books, and scholarly reports related to AI, speech recognition technologies, and pronunciation learning. A qualitative approach is appropriate for this study as it enables an in-depth examination of previous research and provides a deeper understanding of technological and pedagogical developments in language learning. The data were collected from major academic databases, including Google Scholar, Scopus-indexed journals, Web of Science, ERIC, JSTOR, and Science Direct, with priority given to literature published between 2015 and 2026 to ensure recent relevance. The collected literature was analyzed using thematic analysis, focusing on major themes such as pronunciation improvement, feedback effectiveness,

learner motivation, autonomous learning, technological limitations, and pedagogical implications. The identified patterns and recurring findings were systematically examined to develop a comprehensive understanding of the role and effectiveness of AI-based pronunciation learning technologies.

Results

The analysis of the reviewed literature indicates that AI-based speech recognition technologies have significant potential for enhancing pronunciation skills among second language and EFL learners. The findings obtained through thematic analysis reveal several major themes, including pronunciation improvement, effectiveness of feedback, learner autonomy, motivation, personalized learning, and technological challenges. These findings suggest that the integration of Artificial Intelligence into pronunciation learning can create more adaptive, interactive, and learner-centred language learning environments.

Improvement in Pronunciation Skills

The findings show that AI-based speech recognition technologies contribute positively to the improvement of learners' pronunciation abilities. Pronunciation development requires continuous practice, awareness of errors, and corrective feedback. In many traditional language classrooms, providing individual pronunciation guidance is challenging because of limited instructional time and large numbers of students. AI-supported pronunciations tools help reduce this limitation by providing learners with opportunities for independent and repeated practice. AI speech recognition systems analyze learners' spoken language by identifying pronunciation patterns, detecting errors, and providing corrective suggestions. These technologies assist learners in improving both segmental features, such as vowel and consonant sounds, and suprasegmental features, including stress, rhythm, and intonation. The findings support the view of Derwing and Munro (2015), who emphasized that the primary goal of pronunciation learning should be intelligible communication rather than achieving native-like pronunciation. AI pronunciation tools support this objective by helping learners develop clearer and more understandable speech. Through regular interaction with AI-based applications, learners become more conscious of their pronunciation weaknesses and gradually improve their oral communication skills.

Effectiveness of AI-Based Corrective Feedback

Immediate feedback was identified as one of the most important contributions of AI speech recognition technologies. Traditional pronunciation teaching often depends on teacher feedback, which may not always be immediate or individualized. In

contrast, AI systems can instantly evaluate learners' pronunciation performance and provide specific recommendations for improvement. Results indicate that instant corrective feedback helps learners recognize pronunciation errors and make necessary modifications during practice. This process is closely related to Schmidt's (1990) Noticing Hypothesis, which explains that learners need to notice the difference between their own language production and the target language form in order to achieve improvement. AI-based applications provide different forms of feedback, including pronunciation scores, visual representations, audio comparisons, and correction suggestions. These features increase learners' awareness and encourage self-correction. Therefore, speech recognition technologies create an active learning process where students continuously evaluate and improve their own pronunciation.

Learner Autonomy and Independent Practice

Another significant finding is that AI-based pronunciation technologies promote learner autonomy. Unlike traditional classroom learning, AI tools allow students to practice pronunciation without restrictions of time and place. Learners can repeat exercises, monitor their progress, and control their own learning process. The findings correspond with Self-Regulated Learning Theory, which emphasizes the importance of learners' active participation in planning, monitoring, and evaluating their learning. AI pronunciation platforms support self-regulation by providing performance records and personalized learning guidance. According to Warschauer and Healey (1998), technology-assisted language learning increases learners' independence by providing greater access to learning resources. Similarly, AI-based speech recognition systems encourage learners to take responsibility for improving their speaking abilities.

Motivation and Confidence Development

The reviewed studies suggest that AI-based pronunciation tools positively influence learners' motivation and confidence. Many EFL learners experience anxiety when speaking in front of teachers and classmates because they are afraid of making pronunciation mistakes. AI technologies provide a private learning environment where students can practice without social pressure. This comfortable environment reduces speaking anxiety and increases learners' willingness to practice. Repeated pronunciation exercises with AI feedback help students build confidence in their speaking ability. The interactive features of AI applications, such as progress tracking and personalized activities, also increase learner engagement. Godwin-Jones (2018) stated that digital language learning technologies can increase

motivation by creating flexible and interactive learning experiences. Therefore, AI speech recognition tools not only improve pronunciation accuracy but also contribute to learners' psychological readiness for communication.

Personalized and Adaptive Pronunciation Learning

The results reveal that personalization is another important advantage of AI-based speech recognition technology. Every language learner faces different pronunciation problems based on linguistic background, learning experience, and individual ability. Traditional classroom methods may not always address these individual differences effectively. AI systems use advanced algorithms to analyze learners' performance and provide customized exercises according to their needs. This adaptive learning approach allows students to focus on their specific pronunciation difficulties. This finding is related to Vygotsky's (1978) Socio-cultural Theory, particularly the concept of scaffolding. AI tools provide continuous support by guiding learners step by step according to their level of development. As learners improve, the technology adapts learning activities to match their progress.

Challenges and Limitations of AI Speech Recognition Technologies

Although AI-based speech recognition technologies provide many advantages, several challenges were identified. One major issue is the accuracy of speech recognition systems when dealing with different accents and varieties of English. Some AI tools are developed mainly based on native speaker models, which may create difficulties for EFL learners. Accent bias can result in incorrect evaluation of learners' pronunciation. Since English is used globally with many acceptable variations, pronunciation assessment should focus more on intelligibility rather than native-like accuracy (Derwing & Munro, 2015). Another limitation is unequal access to technology. Successful use of AI-based pronunciation tools requires digital devices, internet connection, and technological skills. Learners from economically disadvantaged backgrounds may face difficulties accessing these resources. Privacy is also an important concern because AI speech recognition applications collect and process learners' voice data. Developers and educational institutions should ensure proper ethical standards and data protection measures.

Overall Discussion

The overall findings suggest that AI-based speech recognition technologies can significantly enhance pronunciation learning by providing immediate feedback, personalized instruction, increased practice opportunities, and learner autonomy. These technologies address many limitations of traditional pronunciation teaching

and create more flexible learning experiences. However, AI should not be considered a complete replacement for language teachers. Human instructors remain essential for meaningful communication practice, emotional support, and cultural understanding. The most effective approach is the integration of AI technologies with traditional teaching methods. Therefore, a balanced combination of AI-supported pronunciation training and teacher-guided instruction can create a more effective, interactive, and learner-centred environment for second language learning.

References

1. Ara, R., Chowdhury, M. S. A., & Hasan, M. M. (2026). Teachers' Attitudes toward the Use of Artificial Intelligence in ELT. *Teachers' Attitudes toward the Use of Artificial Intelligence in ELT*, 1(2). <https://doi.org/10.5281/zenodo.21258038>
2. Chapelle, C. A. (2001). *Computer Applications in Second Language Acquisition: Foundations for Teaching, Testing and Research*. Cambridge: Cambridge University Press.
3. Derwing, T. M., & Munro, M. J. (2015). *Pronunciation Fundamentals: Evidence-Based Perspectives for L2 Teaching and Research*. Amsterdam: John Benjamins Publishing Company.
4. Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language Learning & Technology*, 24(1), 16–26.
5. Godwin-Jones, R. (2018). Using mobile technology to develop language skills and cultural understanding. *Language Learning & Technology*, 22(3), 3–17.
6. Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd ed. draft).
7. Krashen, S. D. (1985). *The Input Hypothesis: Issues and Implications*. London: Longman.
8. Levis, J. M. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation*. Cambridge: Cambridge University Press.
9. Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, 32(1), 1–25.
10. Neri, A., Cucchiari, C., & Strik, H. (2008). The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. *ReCALL*, 20(2), 225–243.
11. Pennington, M. C., & Rogerson-Revell, P. (2019). *English Pronunciation Teaching and Research: Contemporary Perspectives*. London: Palgrave Macmillan.

12. Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158.
13. Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
14. Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching*, 31(2), 57–71.
15. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16, Article 39.
16. Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–70.