

# Effectiveness of AI-driven tools in improving English pronunciation training

Jingjing LYU, Hari Krishnan ANDI

Faculty of Social Science, Arts and Humanities, Lincoln University College, Petaling Jaya 47301, Malaysia

**Abstract:** With the rise of globalization, English has become essential for communication worldwide, but pronunciation remains a significant challenge for non-native learners. Traditional training methods often fall short due to delayed feedback and lack of personalized instruction. This study explores the potential of AI-assisted pronunciation training, which utilizes speech recognition, machine learning, and natural language processing to offer personalized, real-time feedback. The research evaluates the effectiveness of AI in improving pronunciation accuracy, fluency, and learner confidence. While AI technology shows promise, challenges like speech recognition accuracy and processing complex speech contexts still need to be addressed. Nonetheless, AI represents a transformative tool for language education.

**Key words:** AI pronunciation training; speech recognition; machine learning; language education; language fluency

## 1 Introduction

As globalization accelerates, English has become a key language for international communication and a core subject in education. However, English pronunciation poses a significant challenge for non-native learners, affecting their communication skills and confidence. Traditional pronunciation training methods often fall short due to delayed feedback and a lack of individualized attention, leaving many learners struggling.

The rapid advancement of artificial intelligence offers promising solutions, particularly through speech recognition, deep learning, and natural language processing. These technologies enable AI to accurately assess pronunciation and provide real-time feedback, overcoming the limitations of traditional methods and enhancing learners' efficiency and accuracy [1].

This study examines the role of AI in English pronunciation training, evaluating its effectiveness in improving accuracy, fluency, and speech recognition. By analyzing AI-assisted tools, the research aims to provide practical and theoretical guidance for future language teaching reforms. AI offers personalized learning experiences, real-time error correction, and objective evaluation, significantly improving pronunciation.

## 2 AI-assisted pronunciation training system design and theoretical framework

The design of an AI-assisted pronunciation training system is grounded in advanced linguistic, speech recognition, and machine learning theories. These technologies work in tandem to offer a highly adaptive and efficient approach to pronunciation training. The first crucial component of the system is phonetic analysis and speech recognition. As the core of AI-based pronunciation training, phonetic analysis enables the system to break down speech into the smallest units of sound—phonemes—and their corresponding syllables. This analysis is crucial because phonemes serve as the foundational

building blocks of language pronunciation. The AI system uses speech recognition algorithms to compare the learner's pronunciation against a reference standard, which allows for an objective evaluation. By accurately identifying specific sounds and comparing them to the ideal pronunciation models, the system provides precise assessments and detailed feedback that is central to improving the learner's spoken English [2].

The second core feature is the integration of machine learning and acoustic modeling. The AI system employs supervised learning algorithms, trained using extensive corpora of native speaker data, to continuously enhance its ability to recognize speech patterns. This training helps the system create an acoustic model that maps sound waves to phonetic symbols. Through this model, the AI system is able to process variations in speech such as accents, speaking rates, and diverse pronunciations, ensuring that the evaluation of speech is not only accurate but also flexible enough to handle different speaking contexts.

Once the learner's speech is processed, the AI compares it with the ideal pronunciation model using an error detection and feedback mechanism. This analysis goes beyond segmental accuracy to evaluate suprasegmental features like intonation, stress patterns, and rhythm. By identifying discrepancies between the learner's speech and the target model, the system provides real-time feedback. This feedback is crucial as it includes not only phonetic annotations but also visual aids, such as spectrograms, and audio comparisons. For instance, the system can highlight areas where stress falls incorrectly or where vowel sounds are mispronounced, allowing the learner to hear the difference and adjust accordingly.

Another strength of AI-assisted pronunciation training is its personalized improvement suggestions. Unlike traditional methods that offer a uniform approach, AI systems adapt to each learner's individual needs. The system continuously analyzes the learner's performance, identifying recurring errors or problem areas — whether it's specific phonemes, stress patterns, or intonation patterns—and customizes the lessons accordingly. If, for example, the system detects that the learner consistently mispronounces the "th" sound, it will prioritize exercises to address that issue in subsequent lessons. This adaptability ensures that training is directly aligned with the learner's specific needs and learning pace, fostering more efficient progress and helping learners build confidence in speaking.

The use of speech recognition, speech synthesis, and natural language processing enhances the system's capabilities. Speech recognition transcribes spoken language into text, allowing the system to compare it with the standard pronunciation model and detect subtle errors in phonemes, syllables, stress, and intonation patterns. It relies on neural networks trained with large native speaker datasets to detect even minor sound quality distinctions. Speech synthesis, or text-to-speech, provides a model pronunciation for learners, offering a clear auditory reference. Learners can compare their attempts with AI-generated speech to help them identify and correct errors in real-time.

Natural language processing is used to assess more advanced pronunciation aspects, such as sentence structure, word stress, and pitch patterns. NLP allows the AI system to process and understand the context in which sounds are used, which is essential for more complex aspects of speech, like intonation and rhythm. For example, the system can determine whether a learner is correctly stressing the syllables in multisyllabic words or whether they are using rising or falling pitch patterns in a sentence. Intonation and rhythm are often neglected in traditional pronunciation training, yet they are critical for clear communication in English. By analyzing these elements, NLP helps the AI system provide feedback that encompasses a broader range of pronunciation features beyond individual sounds.

The pronunciation model development begins with collecting and preprocessing speech data from native speakers, removing noise and enhancing clarity to improve phoneme recognition. After preparing the data, the system is trained using deep learning techniques, such as convolutional neural networks and long short-term memory networks, which are ideal for processing time-dependent speech data. The trained model then maps audio features like frequency patterns and

amplitude variations to corresponding phonetic symbols and words.

### **3 Effectiveness analysis of AI-assisted pronunciation training**

AI-assisted pronunciation training offers key advantages in improving pronunciation accuracy, fluency, and language intuition. By utilizing speech recognition and real-time feedback, AI systems can identify specific pronunciation errors, allowing learners to quickly correct them and refine their speech. The continuous feedback loop accelerates learning by providing immediate corrections, enabling learners to track progress effectively [3].

A major strength of AI-assisted training is its personalized approach. Unlike traditional methods, AI tailors the learning experience to each learner's unique needs, focusing on areas where they struggle, such as specific phonemes or stress patterns. This targeted practice enhances training effectiveness by directly addressing individual weaknesses.

However, there are challenges with AI-assisted pronunciation training. One major issue is the accuracy of speech recognition, particularly when dealing with accents, regional variations, or background noise, which can affect feedback quality. Additionally, AI systems may struggle with complex linguistic contexts, such as varying speech speeds or emotional tones, leading to inaccuracies in pronunciation evaluation.

### **4 Conclusion**

In conclusion, AI-assisted pronunciation training offers significant advantages in improving accuracy, fluency, and personalized learning. It provides quick, targeted feedback, allowing learners to address specific pronunciation issues effectively. However, challenges such as limitations in speech recognition accuracy and difficulties in handling complex speech contexts remain [4]. Despite these obstacles, the potential for AI to enhance language learning is clear, and continued advancements in technology can address these shortcomings, making it a valuable tool for learners.

### **Conflicts of interest**

The author declares no conflicts of interest regarding the publication of this paper.

### **References**

- [1] Lu QJ. 2023. Recognition and evaluation algorithm for English pronunciation syllables based on neural prediction model. *Academic Journal of Computing & Information Science*, 6(10).
- [2] Xu J. 2024. Research-based learning strategies and effectiveness evaluation of English under the assistance of artificial intelligence. *Campus English*, 14: 49-51.
- [3] Abdul MA, Imtiaz H, Illahi B, et al. 2025. Investigating voice onset time in Pakistani English speech. *Speech Communication*, 168.
- [4] Li P, Ning T. 2025. Innovative approaches to English pronunciation instruction in ESL contexts: integration of multi-sensor detection and advanced algorithmic feedback. *Frontiers in Psychology*, 15.