

IMPROVING ENGLISH PRONUNCIATION THROUGH ARTIFICIAL INTELLIGENCE

Muratova Dilorom Mahmudovna

Teacher Uzbekistan State

World languages University

Abstract: The integration of artificial intelligence (AI) into language education has opened new avenues for developing second language pronunciation skills. This systematic review synthesizes findings from 22 empirical studies published between 2016 and 2024 on the use of AI-based pronunciation training tools for English language learners. The review examines the types of AI technologies employed, their effectiveness in improving pronunciation accuracy, and the pedagogical implications for classroom practice.

Keywords: artificial intelligence, pronunciation training, automatic speech recognition, English as a foreign language, computer-assisted pronunciation training

Introduction

Pronunciation constitutes a fundamental component of second language (L2) acquisition, directly impacting learners' communicative competence and overall intelligibility in various communication contexts. For English as a Foreign Language (EFL) learners, pronunciation challenges are multifaceted and often stem from limited access to authentic English-speaking environments and the inherent phonological differences between English and learners' native languages. The difficulties in acquiring accurate pronunciation are further exacerbated by the need for sustained, repetitive practice and immediate feedback—resources frequently lacking in traditional classroom settings due to large class sizes, time constraints, and limited teacher availability.

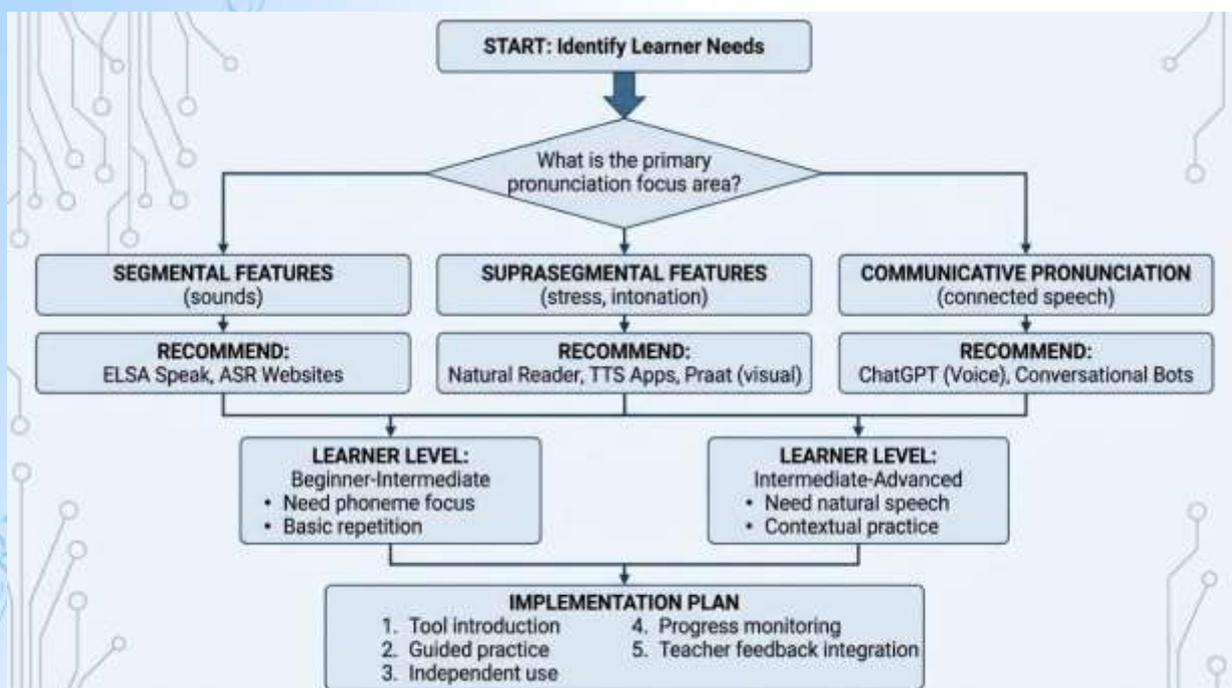
Traditional approaches to pronunciation instruction have relied heavily on teacher-led modeling, listen-and-repeat drills, and minimal pair exercises. While these methods have demonstrated some effectiveness, they are often insufficient for addressing the individualized needs of diverse learners. Consequently, many learners struggle with both segmental features, such as individual vowel and consonant sounds, and suprasegmental aspects, including stress, rhythm, and intonation—elements crucial for fluent and comprehensible communication.

The rapid advancement of information and communication technologies has catalyzed the development of Computer-Assisted Pronunciation Training (CAPT) as an active research

domain . Technology offers a virtual atmosphere that enables learners to interact with native speaker models without the need to travel to English-speaking countries. Through these virtual platforms, learners can simulate pronunciation practice, gaining valuable insights into native-like articulation. Moreover, a plethora of software and applications now exists to assist learners in practicing pronunciation, augmenting learner motivation, reducing anxiety, and providing automated and individualized feedback—a unique feature not easily attainable through face-to-face interactions in natural environments .

More recently, the integration of artificial intelligence has revolutionized language learning by enabling interactive, responsive, and personalized practice . AI-powered conversational agents, such as ChatGPT, and dedicated pronunciation applications, such as ELSA Speak, have shown significant potential in fostering pronunciation skills through immediate, learner-centered feedback. These tools help learners overcome the anxiety often associated with speaking in a foreign language by providing low-pressure practice environments .

Despite the growing body of research supporting the role of AI tools in language learning, the application of AI-driven interventions specifically for enhancing EFL pronunciation skills remains underexplored, particularly in systematically synthesized form. While individual studies have demonstrated positive outcomes, a comprehensive understanding of the range of available tools, their comparative effectiveness, and their pedagogical implications is needed.



AI Tool Selection Flowchart for Educators

Method

This study employed a systematic review methodology to synthesize empirical research on AI-based pronunciation training tools. The review followed established guidelines for systematic reviews in educational research, ensuring transparency, replicability, and rigor in the selection and analysis of studies.

A comprehensive literature search was conducted across multiple electronic databases, including Scopus, Dimensions, SpringerLink, and Google Scholar. The search was limited to peer-reviewed journal articles and conference proceedings published between January 2016 and December 2024. This timeframe was selected to capture the most recent developments in AI technology while ensuring a sufficient body of literature for meaningful synthesis.

The search employed combinations of the following keywords: "artificial intelligence," "AI," "pronunciation training," "pronunciation teaching," "automatic speech recognition," "ASR," "text-to-speech," "TTS," "EFL pronunciation," "computer-assisted pronunciation training," "CAPT," "ELSA Speak," and "ChatGPT pronunciation." Boolean operators (AND, OR) were used to combine search terms appropriately.

Studies were included in this review if they met the following criteria:

1. Population: Participants were English language learners (either EFL or ESL) at any proficiency level.
2. Intervention: The study investigated the use of AI-based tools or technologies specifically for pronunciation training.
3. Outcomes: The study reported empirical data on pronunciation outcomes, learner perceptions, or pedagogical implications.
4. Publication type: The study was published in a peer-reviewed journal or as part of international conference proceedings.
5. Language: The study was published in English.

Studies were excluded if they: (a) focused on AI applications for language skills other than pronunciation (e.g., grammar, vocabulary, writing); (b) were theoretical papers without empirical data; (c) were duplicate publications; or (d) were not available in full text.

The initial search yielded 187 potentially relevant records. After removing duplicates, 156 records remained for title and abstract screening. Two reviewers independently screened titles and abstracts against the inclusion criteria, resulting in 48 studies selected for full-text review. Following full-text assessment, 26 studies were excluded for various reasons (e.g., lack

of empirical data, focus on non-pronunciation outcomes), leaving 22 studies that met all inclusion criteria for final synthesis.

Data were extracted from each included study using a standardized form that captured: author(s) and year, country of study, participant characteristics, AI tool investigated, research design, key findings related to pronunciation outcomes, and reported challenges or limitations.

The methodological quality of included studies was assessed using an adapted version of the Mixed Methods Appraisal Tool (MMAT). Each study was evaluated based on criteria appropriate to its design (e.g., randomization, control groups, valid outcome measures, attrition rates). Studies were not excluded based on quality scores, but quality ratings were considered when interpreting findings.

Results

The 22 included studies were published between 2016 and 2024, with a notable increase in publications after 2022, reflecting growing research interest in AI applications for pronunciation training. The studies were conducted in diverse geographical contexts, including Iran, Oman, Indonesia, South Korea, Taiwan, Turkey, and the United States, demonstrating the global relevance of AI-assisted pronunciation instruction.

Participant populations primarily consisted of tertiary-level EFL learners, though some studies included secondary school students and adult learners in language institutes. Proficiency levels ranged from beginner to advanced, with intermediate learners being the most frequently studied population. Sample sizes varied considerably, from small-scale qualitative studies with fewer than 20 participants to larger experimental studies with over 100 participants.

The reviewed studies employed various AI-based technologies for pronunciation training, which can be categorized into three main types:

Automatic Speech Recognition (ASR) Tools

- ASR technology emerged as the most frequently investigated AI application for pronunciation training. ASR systems evaluate speech input by contrasting it with native speaker models built from databases containing substantial amounts of native speaker speech samples. These systems employ advanced algorithms to convert spoken words into text while simultaneously analyzing pronunciation accuracy.

- ELSA Speak (English Language Speech Assistant) was the most commonly studied ASR-based application. This mobile application uses AI to provide detailed feedback on pronunciation at the phoneme level, highlighting specific sounds that require improvement and offering targeted practice activities. Studies by Anggraini (2022), Kholis (2021), and Ngoc and

Thanh (2024) demonstrated that ELSA Speak significantly improved learners' pronunciation accuracy, particularly for segmental features such as individual vowel and consonant sounds .

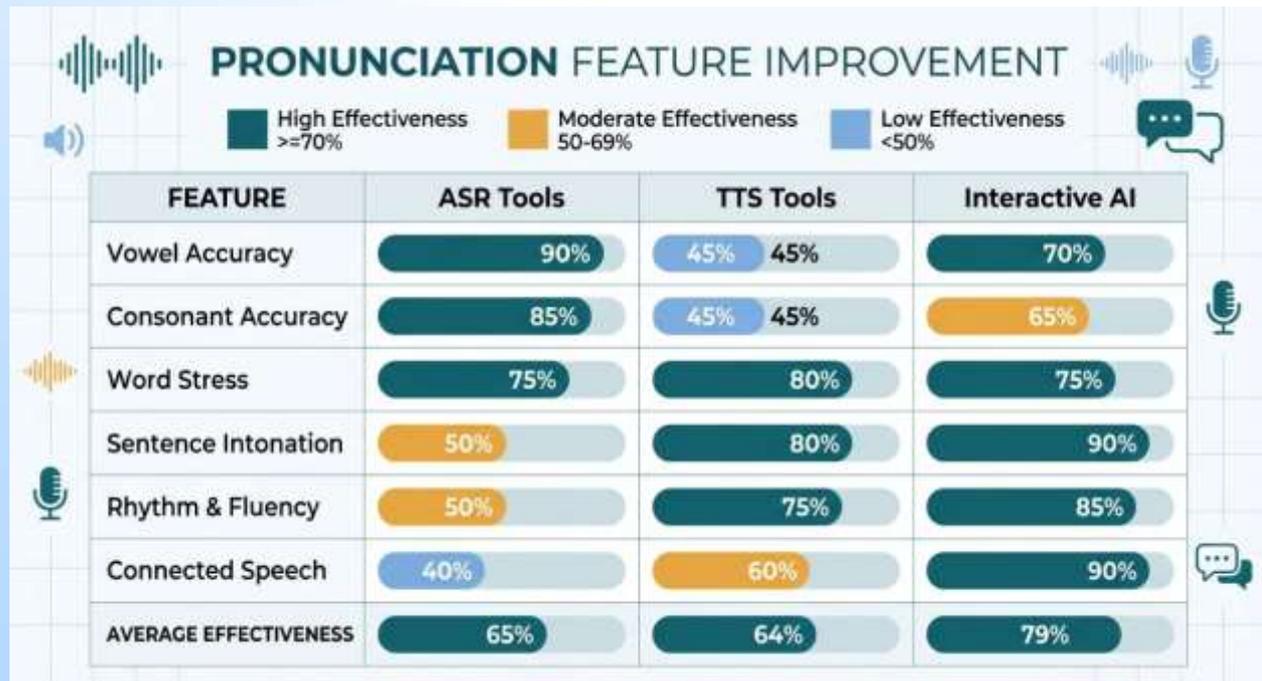
• Other ASR-based tools investigated included mobile dictation applications and web-based ASR systems. Bashori et al. (2022, 2024) examined ASR-equipped websites for pronunciation and vocabulary learning, finding that learners benefited from the immediate feedback and autonomous practice opportunities these platforms provided .

Discussion

This systematic review synthesized findings from 22 empirical studies on AI-based pronunciation training tools for English language learners. The results demonstrate that AI technologies—particularly ASR applications, TTS tools, and interactive conversational agents—hold significant promise for enhancing pronunciation instruction. The findings align with and extend previous research on CAPT, revealing the unique affordances of AI-powered tools for providing immediate, individualized, and low-anxiety pronunciation practice.

The effectiveness of AI-assisted pronunciation training can be understood through multiple theoretical lenses. From a cognitive perspective, the phonological loop component of working memory—responsible for temporary storage and processing of verbal information—plays a crucial role in pronunciation learning . AI tools support the rehearsal loop by providing repeated exposure to accurate pronunciations and immediate opportunities for practice and correction.

From a sociocultural perspective, AI tools can be viewed as mediating artifacts that scaffold learners' pronunciation development. Vygotsky's zone of proximal development is relevant here: AI tools provide support that enables learners to perform beyond their current independent capabilities, with the level of scaffolding adjusting to individual needs.



Conclusion

This systematic review provides evidence that AI-based pronunciation training tools can significantly enhance English language learners' pronunciation accuracy while reducing speaking anxiety and promoting autonomous practice. ASR applications such as ELSA Speak, TTS technologies, and interactive conversational agents like ChatGPT offer immediate, individualized feedback on both segmental and suprasegmental features of pronunciation—capabilities that address longstanding challenges in pronunciation instruction.

As AI technologies continue to evolve, their potential for pronunciation training will likely expand. The challenge for educators and researchers is to harness this potential while maintaining pedagogical integrity and ensuring equitable access to these powerful learning tools. By addressing the gaps identified in this review—particularly regarding teacher education, long-term outcomes, and integration strategies—the field can move toward evidence-based implementation that optimizes AI's contribution to pronunciation learning.

References

1. AbuSeileek, A. F. (2007). Computer-assisted pronunciation instruction: Using Mouton Interactive Introduction to Phonetics and Phonology. *Teaching English with Technology*, 7(3).
2. Al-Jarf, R. (2022). Text-to-speech software for promoting EFL freshman students' decoding skills and pronunciation accuracy. *Journal of Computer Science and Technology Studies*, 4(2), 19-30.
3. Anggraini. (2022). Improving students' pronunciation skill using Elsa Speak application. *Journey: Journal of English Language and Pedagogy*, 5(1), 135-141.

4. Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2022). 'Look, I can speak correctly': Learning vocabulary and pronunciation through websites equipped with automatic speech recognition technology. *Computer Assisted Language Learning*, 37(5-6), 1335-1363.
5. Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024). I Can Speak: Improving English pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching*, 2, 1-19.
6. Dai, Y., & Wu, Z. (2023). Mobile-based dictation ASR for pronunciation training: A systematic review. *Computer Assisted Language Learning*.
7. Dida, H. A., Chakravarthy, D., & Rabbi, F. (2023). ChatGPT and big data: Enhancing text-to-speech conversion. *Mesopotamian Journal of Big Data*, 2023, 31-35.
8. Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Exploring the impacts of an AI-driven instructional intervention on Iranian EFL learners' pronunciation skill development. *Discover Education*, 4(3), 307.
9. Fitria, T. N. (2022). Utilizing text-to-speech technology: Natural Reader in teaching pronunciation. *JETLEE: Journal of English Language Teaching, Linguistics, and Literature*, 2(2), 70-78.
10. Gorjian, B., Hayati, A., & Barouni, P. (2013). The effects of computer-assisted pronunciation teaching on EFL learners' use of suprasegmental features. *Journal of Teaching Language Skills*, 5(2), 93-118.