European International Journal of Philological Sciences

TYPE Original Research
PAGE NO. 53-57
DOI 10.55640/eijps-05-12-15



# AI Tools For Speaking Fluency And Pronunciation: Effectiveness And Limitations

Babayeva Komila Rishatovna

Department of Foreign Languages Education at Tashkent State University of Economics, Uzbekistan

OPEN ACCESS

SUBMITED 12 October 2025
ACCEPTED 20 November 2025
PUBLISHED 17 December 2025
VOLUME Vol.05 Issue 12 2025



**Abstract**: Artificial intelligence (AI) is changing how we teach speaking in a second language (L2) by making feedback on pronunciation and interactive speaking practice available outside of the classroom. There are two main types of tools that are used today: automatic speech recognition (ASR) systems that turn speech into text and give corrective signals, and conversational agents (like chatbots and speech-enabled assistants) that mimic conversation and keep people talking. Research increasingly indicates that these tools can facilitate quantifiable improvements in specific facets of pronunciation and speaking performance, especially when learners are provided with frequent practice opportunities and when AI-mediated feedback is augmented by peer or teacher support. Classroom studies show that ASR-supported practice can help students' pronunciation and speaking test scores in some situations compared to traditional teaching. It can also boost students' confidence and willingness to speak. There are still some big problems, though. ASR doesn't always rate speech with accents in ways that match how people think it sounds, and its feedback can change depending on the speaker, the task, and the phonological target. Conversational agents can boost the number of interactions, but they might not show how complicated the language is in real life and might favor speech that is easy for machines to understand over speech that is easy for people to understand. Emerging syntheses also underscore ethical risks (privacy, data retention, bias) and pedagogical risks (overreliance, diminished learner autonomy if feedback is not explicitly taught). This article examines research trends, suggests a pragmatic approach for assessing AI tools for speaking fluency and pronunciation, and delineates the circumstances in which advantages are

most likely to translate to actual communicative competence.

**Keywords:** AI, automatic speech recognition, learning how to pronounce words, speaking fluency, chatbots, EFL, corrective feedback, and intelligibility.

**Introduction:** Being able to speak well is a cognitive, linguistic, and social skill all at once. Fluency requires real-time processing, which includes planning ideas, picking words, putting together grammatical structures, and speaking while also keeping an eye on how well others understand you and interacting with them in the right way. Fluency and pronunciation go hand in hand. Segmental accuracy (vowels and consonants), suprasegmental control (stress, rhythm, intonation), and timing (pausing, speech rate) all affect how easy it is to understand and how hard it is for the listener. In many English as a Foreign Language (EFL) settings, students don't get enough time to speak or get regular feedback, especially on their pronunciation. There isn't enough time in class, the teacher's attention is divided, and students don't always get along with each other. Because of these structural limits, speaking skills are a good candidate for help from technology.

AI tools offer two benefits that traditional computer-assisted language learning (CALL) has had trouble scaling: frequent personalized feedback and interactive practice on demand. ASR-based applications put a practical principle into action: if a system can accurately transcribe a learner's speech, the learner can assume that their production is closer to an intelligible target. But ASR is not just a neutral way to measure. It incorporates statistical assumptions, training data distributions, and optimization objectives that may not align with the communicative aims of L2 instruction. Recent research in language instruction has transitioned from inquiring whether ASR "functions" to investigating what it specifically measures and how its signals can be synchronized with pronunciation pedagogy.

At the same time, conversational agents have gone from scripted chatbots to systems that can create dialogue that is relevant to the situation. The main idea behind teaching speaking is "time on task": students who talk more about different subjects and with less anxiety should get better. Educational reviews of AI chatbots mention benefits that come up again and again, like the ability to personalize and more chances to practice. They also talk about problems with reliability, accuracy, and ethical issues. But when it comes to speaking practice, "more is better" isn't always true. The quality of interaction is important: feedback must be clear, task requirements must be close to real-life communication, and learners must learn forms instead of just adapting to the limits of the tools.

Empirical studies illustrate both potential and risk. For example, a mixed-method classroom investigation with intermediate EFL learners reported that an ASR-plus-peer-correction approach outperformed teacher-led feedback in several speaking-related measures, including performance on speaking assessments. Such findings support investment in AI-mediated speaking practice, but they do not settle the question of transfer: do gains on tool-aligned tasks generalize to spontaneous interaction with human listeners? Research comparing ASR recognition to human intelligibility judgments indicates that alignment can vary substantially by speaker and task type, raising concerns about using ASR scores as a stand-alone proxy for communicative success.

Given these tensions, stakeholders need a rigorous methodology for evaluating AI speaking tools—one that integrates outcomes (fluency, pronunciation, confidence), measurement validity (human and machine rating alignment), and contextual factors (learner level, instructional design, language background). This article addresses that need by synthesizing current evidence, articulating a defensible evaluation methodology, and offering a balanced interpretation of effectiveness and limitations.

The aim of this article is to analyze the effectiveness and limitations of AI tools for developing speaking fluency and pronunciation in EFL/L2 settings and to propose a practical, research-informed methodology for evaluating such tools in instructional and institutional contexts. The article emphasizes communicative validity, transfer to human interaction, and responsible deployment.

This article adopts a structured narrative review approach combined with a methodological framework proposal. The reviewed evidence is anchored in peer-reviewed research on ASR-based pronunciation practice, learner perceptions of ASR applications, and AI-mediated interactive speaking tasks, complemented by broader systematic reviews of educational chatbots.

The methodological framework proposed here operationalizes evaluation across three levels: (1) tool-level performance, (2) learning-level outcomes, and (3) transfer-level outcomes. Tool-level performance includes the stability of ASR transcription and scoring under accented speech conditions and the transparency of feedback. Learning-level outcomes include changes in pronunciation and speaking proficiency as measured

by established speaking tasks and by complexity–accuracy–fluency (CAF) metrics when appropriate. Evidence supporting CAF-based speaking evaluation in AI-mediated game contexts suggests that oral proficiency changes can be quantified beyond simple "scores," enabling more diagnostic interpretation of gains. Transfer-level outcomes refer to improvements that persist outside the tool environment, such as better intelligibility for human listeners and improved performance on standardized speaking tests.

For pronunciation, the framework prioritizes intelligibility and comprehensibility over native-like accent reduction, consistent with contemporary pronunciation pedagogy and with concerns that ASR may reward "system-optimized" speech. The framework therefore recommends dual assessment: human ratings (intelligibility/comprehensibility and global speaking performance) alongside ASR-derived indicators (word error rate, phoneme-level diagnostics when available, and consistency across prompts). The need for such dual assessment is supported by evidence that ASR and human listeners may identify similar error types yet diverge in speaker-level correlations, depending on the task and individual learner profiles.

For fluency, the framework focuses on temporal measures such as speech rate, mean length of run, and pausing behavior, and it treats affective variables (anxiety, willingness to communicate, self-efficacy) as mediators rather than peripheral outcomes. Research linking AI-supported speaking practice to reduced speaking anxiety provides justification for including these mediators in evaluation, because fluency in real interaction is constrained by both competence and affective readiness.

Finally, the framework integrates a usage analytics component, recognizing that learning effects depend on how learners actually engage with tools over time. Evidence from observational work on autonomous ASR-based pronunciation practice indicates that learner behavior across sessions is variable and therefore must be monitored to interpret outcomes and to design effective scaffolding.

Evidence to date supports a cautiously positive interpretation of AI tools for speaking development, with the strongest results occurring when tools are embedded in pedagogically coherent designs rather than used as isolated self-study add-ons. In classroom contexts, ASR-supported practice can improve pronunciation-related outcomes and speaking test performance, particularly when learners practice frequently and receive additional social support through peer correction. The mixed-method study by Sun, for example, used read-aloud tasks, spontaneous conversations, and an IELTS speaking test to evaluate outcomes and reported advantages for the ASR plus peer-correction condition over teacher-led feedback. The practical implication is not that AI replaces teacher feedback, but that it can increase the volume and immediacy of feedback signals, freeing classroom time for higher-order communicative work.

Pronunciation benefits, however, appear to be more reliable for certain dimensions than others. ASR tools are generally well suited to segmental practice and to repeated production of short utterances, where learners can iteratively modify output and observe changes in transcription accuracy. This aligns with language-teaching analyses emphasizing ASR's potential as a practice engine for high-frequency production and noticing. Yet the same analyses highlight a central constraint: ASR feedback is often indirect (through transcription) and may not specify the phonetic source of error, especially for suprasegmentals. Consequently, learners may improve on items that the system is sensitive to, while prosodic naturalness and pragmatic appropriateness remain undertrained.

Measurement validity is a pivotal limitation. The study comparing native listener intelligibility judgments to ASR recognition found that despite similar aggregate recognition scores, speaker-level alignment varied substantially; ASR mirrored human judgments well for some speakers and much less for others, implying that ASR-based scores cannot be assumed to generalize across individuals. This matters operationally: if institutions use ASR scores for placement, progress, or high-stakes decisions, they risk misclassifying learners whose accent patterns are underrepresented in ASR training data or whose speech features interact poorly with the system's acoustic and language models.

Usage behavior further complicates effectiveness claims. Learners do not necessarily use tools in the ways designers intend. Monitoring research on ASR-based autonomous pronunciation practice highlights that engagement patterns can fluctuate across sessions and that such variation is consequential for outcomes. Some learners may focus narrowly on "getting the system to recognize" rather than building robust phonological categories. Others may overpractice a small set of items, generating local gains without broader transfer. These patterns justify integrating learning analytics and explicit training in how to interpret and act on feedback, especially in self-access environments.

Conversational agents offer a different value proposition: they can sustain interaction and reduce barriers to speaking by providing non-judgmental

practice partners. Broader systematic reviews of AI chatbots in education suggest that learners benefit from personalization and skill development opportunities, while educators benefit from time-saving support; at the same time, reliability and ethical concerns remain prominent. In EFL speaking practice specifically, a systematic review of AI-powered chatbots for speaking has been reported to identify benefits such as real-time conversation practice and individualized support, but such syntheses also imply uneven evidence quality and substantial heterogeneity across tools and outcome measures. This heterogeneity limits confident generalization and underscores the need for standardized evaluation methodology.

Studies comparing interaction with AI versus human partners illuminate a key limitation: AI can change what "successful communication" looks like. In a game-based speaking study using CAF metrics, learners interacted with either AI or human partners, illustrating that task conditions can shape complexity, accuracy, and fluency outcomes in different ways. Qualitative learner feedback from such work suggests that AI may be more convenient and predictable, but also less tolerant of "rough" language and less capable of negotiating meaning in human-like ways. This tension creates a risk that learners optimize for machine interpretability, which is not identical to human comprehensibility. From a pedagogy perspective, the remedy is to frame AI as a rehearsal space and diagnostic tool, while retaining human interaction as the benchmark for communicative adequacy.

Affective outcomes are among the most practically important, even when linguistic gains are modest. Speaking anxiety suppresses participation and reduces opportunities for feedback and interaction. Evidence that AI-supported speaking practice can reduce speaking anxiety and improve speaking performance supports the inclusion of affective variables in evaluation and justifies AI adoption as an access and participation intervention. Nonetheless, anxiety reduction should not be conflated with proficiency growth; it is best interpreted as a mediator that can enable more practice, which then may lead to linguistic development if tasks and feedback are well designed.

Finally, ethical and governance issues are not peripheral—they can determine whether AI tools are sustainable in institutional settings. Systematic reviews of AI chatbots emphasize concerns about accuracy, reliability, and ethical considerations. For speaking tools, privacy risks are amplified because voice data can be personally identifying. Institutions therefore need procurement and deployment policies that address data minimization, retention, and transparency. Pedagogically, academic integrity concerns also arise when conversational agents generate language for learners; for speaking practice, the main risk is not plagiarism but dependency and reduced self-generated production if learners shift from producing speech to curating AI output.

Overall, the evidence indicates that AI tools are most effective when they are deployed as part of a blended speaking methodology: ASR supports high-frequency controlled practice and feedback cycles, while conversational agents increase speaking time and lower affective barriers. The major limitations—measurement validity, accent bias, weak suprasegmental feedback, and ethical risks—are manageable when evaluation is dual-mode (human plus machine), when learners are taught feedback literacy, and when institutions adopt explicit data governance.

AI tools for speaking fluency and pronunciation can deliver meaningful benefits, primarily by expanding opportunities for practice and by providing rapid feedback that supports noticing and iterative improvement. Classroom evidence shows that ASR-integrated approaches, particularly when combined with peer correction, can improve pronunciation and speaking assessment outcomes relative to traditional feedback alone. At the same time, ASR scores do not consistently align with human judgments of intelligibility across speakers and tasks, which limits their validity as stand-alone indicators of communicative success. Conversational agents can increase interaction volume and confidence, but they may shape learner behavior toward "machine-optimized" speech and may underrepresent pragmatic complexity, requiring careful task design and teacher mediation.

A defensible methodology for adoption and evaluation should therefore include dual assessment (human and ASR-based measures), validated speaking tasks (including spontaneous interaction), usage analytics to interpret engagement patterns, and explicit instruction in how to interpret AI feedback. Monitoring learner behavior is not optional: it is necessary to understand whether outcomes reflect genuine learning or strategic adaptation to the tool. Responsible deployment also requires institutional governance addressing privacy, transparency, and ethical risks, consistent with the broader literature on AI chatbots in education. With these safeguards and design principles, AI can function as a scalable enhancement to speaking pedagogy rather than a fragile substitute for human interaction.

## REFERENCES

**1.** Sun W. The impact of automatic speech recognition

technology on second language pronunciation and speaking skills of EFL learners: a mixed methods investigation // Frontiers in Psychology. 2023. Vol. 14. Art. 1210187. DOI: 10.3389/fpsyg.2023.1210187.

2. McCrocklin S., Levis J. Automatic speech recognition and pronunciation learning // Language Teaching. 2025. Published online: 06 Aug 2025. DOI: 10.1017/S0261444825100852.
3. Inceoglu S., Chen W.-H., Lim H. Assessment of L2 intelligibility: Comparing L1 listeners and automatic speech recognition // ReCALL. 2023. Vol. 35, No. 1. P. 89–104. DOI: 10.1017/S0958344022000192.
4. Inceoglu S., Chen W.-H., Lim H. Monitoring learners' behavior during ASR-based pronunciation practice // System. 2024. Vol. 124. DOI: 10.1016/j.system.2024.103387.
5. Pituxcoosuvarn M., Tanimura M., Murakami Y., White J. S. Enhancing EFL Speaking Skills with AI-Powered Word Guessing: A Comparison of Human and AI Partners // Information. 2025. Vol. 16, No. 6. Art. 427. DOI: 10.3390/info16060427.
6. Farrús M., Cuadros M., Diab M., Griol D. Intelligent Computer-Assisted Language Learning for Pronunciation and Oral Proficiency: A Systematic Review // Applied Sciences. 2023. Vol. 13. Art. 9399. DOI: 10.3390/app13169399.
7. Moustafa A. M., et al. When artificial intelligence meets pronunciation: exploring learners' perceptions of ASR apps in L2 pronunciation learning // ReCALL. 2024. Published online: 20 May 2024. DOI: 10.1017/S0958344024000015.
8. Labadze L., Grigolia M., Machaidze L. Role of AI chatbots in education: systematic literature review // International Journal of Educational Technology in Higher Education. 2023. Vol. 20. DOI: 10.1186/s41239-023-00426-1.
9. Du J., Daniel B. K. Transforming Language Education: A Systematic Review of AI-Powered Chatbots in EFL Speaking Practice // Computers and Education: Artificial Intelligence. 2024. Art. 100230. DOI: 10.1016/j.caeai.2024.100230.
10. Fryer L. K., Coniam D., Carpenter R., Lăpușneanu D. Bots for language learning now: Current and future directions // Language Learning & Technology. 2020. Vol. 24, No. 2. P. 8–22. URL: http://hdl.handle.net/10125/44719 (accessed: 14.12.2025).
11. El Shazly R. Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study // Expert Systems. 2021. Vol. 38, No. 3. Art. e12667. DOI: 10.1111/exsy.12667.
12. Godwin-Jones R. Chatbots in language learning: AI systems on the rise // In: Intelligent CALL, granular systems and learner data: short papers from EUROCALL 2022. 2022. URL: https://files.eric.ed.gov/fulltext/ED625175.pdf (accessed: 14.12.2025).