NUML Journal of Critical Inquiry
Vol. 23 (II) December 2025

e ISSN 2789-4665; p ISSN 2222-570
https://doi.org/ 10.52015/numljci.v23iII.316

Adopting AI-Based Speech Recognition for Enhancing Pronunciation Accuracy and Learner Confidence

Rehana gulzar[1] & Muhammad Ehsan[2]

ABSTRACT

Article History:

Received:
October 3, 2025
Accepted:
December 23, 2025

Funding:
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest:
The authors have declared no potential conflicts of interest and falsification/fabrication of data with respect to the research, authorship, and/or publication of this article.

This study examines the integration of AI–based speech recognition (ELSA Speak, Speechling) in a language centre and its effects on pronunciation accuracy and learner confidence. Two groups of intermediate EFL learners (N = 40) completed pre-test and post-test diagnostic pronunciation assessments, with the experimental group receiving support through AI-app usage analytics. The comparative results revealed that the AI-assisted group demonstrated significantly more positive outcomes than the control group ($p < .001$), thus achieving higher post-test scores and increased learner confidence in the pronunciation assessment in the experimental group (paired t, $p < .001$). Consistent engagement in AI-assisted pronunciation practice was associated with improved performance ($r = .62$, $p < .01$). Quantitative results showed that learners using AI tools achieved higher pronunciation scores and confidence, while qualitative findings indicated reduced speaking anxiety, sustained motivation, and enhanced learner autonomy, despite minor implementation challenges such as connectivity issues and occasional feedback perception mismatches. The findings show the effective integration of AI-based pronunciation tools with conventional language instruction through feedback and self-regulated learning scaffolds. However, the intact–class design, single-centre context, and short duration result in limiting generalizability, demanding a longitudinal, multi-site replication.

Keywords: *AI-based speech recognition; pronunciation accuracy; learner confidence; actionable feedback; self-regulated learning*

[1] Lecturer in the Department of English at the International Islamic University, Islamabad. She may be accessed at rehana.gulzar@iiu.edu.pk

[2] Professor at the Centro de Bachillerato Tecnológico Agropecuario No. 162, Francisco I. Madero, Tlaxcala, Mexico. He may be accessed at Muhammadehsan2000@yahoo.com

This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0)

Introduction

The landscape of language education is undergoing significant advancement with transformative tools of Artificial Intelligence (AI). With the emergence of innovative technologies, guidelines for improving speaking and pronunciation have expanded due to helpful speech recognition software, such as ELSA Speak and Speechling. It accommodates immediate response to pronunciation problems with its feature of automated feedback, changing the course of language learning with improved capability to counter conventional hurdles of perception, articulation, and limited time in the formal classroom environment.

The advancements in Computer-Assisted Language Learning (CALL) due to Artificial Intelligence (AI) based speech software have transformed the pathway of pronunciation education by addressing learners' problems, such as the anxiety of public error correction and limited learning time. Given the immediate feedback responses on segmental targets (vowels, consonants) and suprasegmental control (stress, rhythm, intonation) yielded by AI-powered software such as ELSA Speak and Speechling, these tools contribute substantially to achieving learning outcomes (Tran & Vu, 2024). Furthermore, the software integrates theory and balanced instructional channels (e.g., technology acceptance, self-regulated learning, and cognitive effective models of multimedia learning) with effective interactions, leading to significant patterns in the domain of intelligence and confidence, with innovative pedagogical layouts and pre-existing second language acquisition (SLA) mechanisms (Dja'far & Hamidah,2024; Kawaguchi,2021; Lantolf & Xi, 2023; Li & Lan, 2022). However, empirical inquiries (feedback design, practice mechanism, communicative patterns) suggest greater durability and command of communication with effective access and automated results.

The interplay between the pronunciation correctness and learner affective variables (confidence, anxiety, and motivation) has been recognised by traditional second language acquisition research methodologies as well as general language educational practices (Pennington & Rogerson-Revell, 2019; Ma et al., 2024). The factors responsible for hampering a learner's progress mainly arise from their deep-rooted fear of negative feedback and the limited number of verbal activities. Research studies have consistently reported the effectiveness of ELSA Speak and highlighted learners' positive perceptions of its use in pronunciation instruction. The evidence of pronunciation improvement was explored by Dja'far and Hamidah (2024), who elucidated that its effectiveness was attributed to its essentials, specifically the layout of display content, gamification aspects, and learner responses. Similarly, gains were recorded in pronunciation accuracy when compared with the control group in the quasi-experimental study using the ELSA Speak sample over a time frame (pre-test-post-test).

In another study, Rahayuningsih, Hartono, & Wahyuni (2025) reported high satisfaction levels with ELSA's speaking pronunciation courses. Among the most salient were perceived ease and confirmation (i.e., alignment between expectations and experience), influencing learners' evaluation of the tool. Adding to this notion, another analysis was carried out to explore those students who spoke in favour of ELSA Speak and its scaffolding, as well as pronunciation practice sessions and course integration. Gusrianto (2025) asserted instructors' role in adaptations and motivational affordances (gamified prompts) advancing towards an interactive experience. The present research study focuses on the recognition and inclusion of AI-mediated speech in educational settings to promote the integration of conventional methods in English spoken classes for ultimately improved pronunciation accuracy.

Literature Review

Research on AI-mediated pronunciation suggests that automatic speech recognition (ASR) can meaningfully complement classroom instruction by providing immediate, individualised feedback in a low–stakes practice space. Ping and Tao (2025) highlighted that across classroom and out-of-class implementations, learners typically reported that the ability to rehearse privately, receive instant cues, and repeated practice reduced the social costs of speaking practice sessions and sustained engagement over time. This aligns with long-standing arguments in second language acquisition, which suggest that access to timely, specific feedback and opportunities for productive output are crucial for phonological development.

A research study conducted by Mohammad Karimi (2024) found that automatic speech recognition (ASR) provides meaningful support for enhancing L2 pronunciation with explicit feedback and effective task designs through meaningful engagement for authentic production. Ngo et al. (2024) carried out a meta-analysis of ESL/EFL studies' reports, which highlighted the positive outcomes of ASR-supported training and the impact of corrective feedback measures aimed at phoneme-level cues, stress patterns, and intonation in digital environments. It was observed that peer scaffolding in the classroom environment helped improve pronunciation and also reduced anxiety and hesitation during communicative interactions with peers in class. In another study, a multi–week ASR program demonstrated lower anxiety levels and higher confidence and motivation in different practice activities in class with effective teacher mediation to overcome speaking issues in spontaneous tasks with clear, actionable cues at the level of individual phonemes, syllable stress, and intonation patterns (Abimanto & Sumarsono, 2024). Denies (2024) observed that when examples were provided (model audio, visualisations of stress or pitch, articulatory hints), learners more easily connected feedback to concrete revisions in their production. Systems that

enabled quick, guided retries appeared to foster noticing and the iterative refinement of difficult phonetic contrasts.

Zou et al.'s (2021) work study the effects of motivation on pronunciation highlighted the role of low-pressure practice, gamified progression, and visible mastery paths. Learners often described ASR environments as less anxiety-provoking than public correction, which, in turn, was associated with a greater willingness to communicate and more persistent practice (Aryanti & Santosa, 2024). Short, bite–sized activities and micro–goals were repeatedly identified as important design features that help learners maintain momentum alongside regular coursework (Syed, 2015). Mubarok & Aziez (2024) highlighted the importance of the language teachers' immediate and constructive feedback. When teachers adjusted targets, modelled how to interpret app feedback, and integrated ASR tasks with weekly phonological aims, gains tended to be more coherent and transferable to communicative work. Conversely, fully autonomous use drifted toward "safe" items in the absence of systematic progression, which suggested that instructor mediation, task variety, and explicit monitoring routines remained pedagogically valuable.

Evidence specific to ELSA Speak and Speechling suggested broadly positive learner perceptions of usability and usefulness, particularly when content sequencing, gamification, and individualised feedback were leveraged to scaffold practice (Dai & Wu, 2023). Adoption and sustained use were strengthened when learners found the interfaces intuitive and when app evaluations aligned with their own perceptions of improvement. At the same time, instructors noted that occasional divergences between automated scoring and learners' intuitions required clarification and classroom discussion.

Mahmood & Bui (2024) emphasised the importance of assessment orientations that prioritise intelligibility and comprehensibility over a narrow pursuit of nativeness. This shift was echoed in teacher reports, where the most valued outcomes were clearer segmental contrasts and more consistent control of stress and rhythm in connected speech (Georgiou, 2025). From this perspective, ASR was found most effective when it helped learners become more understandable to a wider range of interlocutors and when classroom assessment criteria aligned with this objective. Pham & Pham (2025) conducted their research on EFL university students who demonstrated remarkable progress in their speaking and pronunciation using ELSA Speak. They found high satisfaction with the useful approaches and strategies provided by AI tools for enhancing speaking skills. The sustained practice with ease and immediate feedback brought prompt responses in meaningful contexts. They highlighted the role of the teacher who could work on feedback for improving weak areas like pitch variations, stress, and intonation practice for achieving intelligibility-oriented learning goals.

A research study conducted by Karim et al. (2023) examined the implications of the linguistic trends for enhanced speaking skills with the intervention of AI tools, which were curricula embedded as a component of pedagogic practices. Tran & Vu (2024) investigated how EFL students could engage effectively with AI tools to achieve desired outcomes. They examined how variations occurred in different contextual settings. The clustered usage patterns demonstrated the speech variations, which were measured on steady, regular, and irregular scales to interpret spontaneous linguistic behaviours. They found intensive practice also improved their confidence and motivation levels. There were distinct rapid improvements that were marked as possible due to micro-practice sessions in language classes. Students consistently faced stress-related issues and anxiety in public speech. These issues were systematically studied for effective practical solutions. These issues were addressed with instructional plans and concrete strategic measurements in line with the digital spaces. AI-powered speech recognition tools aimed to mitigate these constraints by providing learners with a safe practice environment, where they could repeat utterances and receive real-time corrective feedback. They aimed to provide pronunciation practice where there was no stress and a live audience or listeners who would judge their manner of speech (Vančová, 2023). Recent empirical and meta-analytic work provided growing support for the efficacy of such technologies, enhancing both pronunciation accuracy and learners' confidence and satisfaction (Aryanti & Santosa, 2024; Azam, 2012; Caiza, Villafuerte, & Guanche, 2025).

The AI tools have been created for practicing pronunciation in varied contextual settings. For example, a meta-analysis of 15 automatic speech recognition (ASR) studies on ESL and EFL pronunciation found a medium overall effect size (g=0.69). For pronunciation accuracy, different contextual settings are not a variable anymore, due to the creation of AI speech software. A significant improvement was exhibited in their pronunciation accuracy, as marked review pointing towards the pedagogical value of properly established ASR systems was provided (Abdelhalim & Alsehibany, 2025). As reported by Dai & Wu (2023), the conduction of six autonomous pronunciation sessions with 46 EFL learners revealed the variable engagement of learners with safe items and review. Furthermore, the key findings reflected how few learners strived for difficult articulations based on their communicative patterns and interactive sessions. The feedback interface and provision of corrective cues helped resolve infrastructural issues. In targeting uptake of feedback, strategies used by the learners were important.

The research studies constantly brought attention to implement conditions as the main deciding factors of the result (Azam, 2011; Fountoulakis, 2024; Sari, 2023). For improved output, infrastructure elements such as reliable connectivity,

equitable device access, and brief training were prerequisites. Besides this, other practices proved beneficial for translating automated scores to durable change, like self-paced learning sets, goal setting, progress monitoring, and quick reflection, with remarkable results on efficient practices. Instructors with the usage of a learning analytics dashboard could encounter repeatedly occurring segmental and suprasegmental problems proficiently. It becomes crucial to keep the robustness of gains and clearer reporting on the variation of feedback across fluency bands. The timely and underexplored recognition has been addressed in contemporary research for the progress of pronunciation accuracy and learner confidence with AI as systematic empirical work force bridging the gaps of technological innovation, pronunciation pedagogy, and learning. The present research can contribute to effective results in both theoretical and practical domains, integrating language instructor methodologies with new advanced AI tools.

The following research questions are likely to be helpful in navigating through this paper.

RQ1: To what extent does the use of AI-based speech recognition software improve the pronunciation accuracy of ESL learners compared to traditional classroom practice?

RQ2: How does the adoption of AI-powered pronunciation tools influence learners' confidence and willingness to communicate in English?

RQ3: What are learners' perceptions and attitudes toward the usability, feedback quality, and motivational features of speech recognition applications such as ELSA Speak and Speechling?

Methodology

This study adopted a quasi-experimental research design within the institutional setting of the English Language Learning Center in Islamabad, Pakistan. The design was chosen to capture integration of AI-based speech recognition applications, like ELSA Speak and Speechling. These AI tools had been incorporated into the language support program to examine their impact on the learners` pronunciation and accuracy, and to build their confidence in authentic speaking environments. The approach combined quantitative assessments of learner performance with qualitative insights from classroom observations, learner reflections, and research context, as well as participant feedback. The research was conducted in one of the English language centers in Islamabad, which provided English language support to undergraduate and postgraduate students.

The present study consisted of 40 learners aged 18-24 who were enrolled in an intermediate-level course, titled 'English Pronunciation and Speaking Skill Course'. Participants were divided into two intact groups.

Experimental Group (n = 20): This group integrated AI-based speech recognition tools. ELSA Speak and Speechling as a supplementary practice alongside regular instruction.

Control Group (n=20): This group studied with the traditional methodology, which included phonetic drill practice, minimal pair practice, and teachers` feedback without AI.

All participants had previously completed a placement test at the center, ensuring a relatively homogeneous proficiency level (CEFR B1-B2). There was a pronunciation diagnostic test at the beginning of the course. All the learners completed the standardized diagnostic test, which included reading aloud a word list, sentence structures, and a short passage. It was recorded for later assessment to check segmental accuracy, including consonant and vowel production, and suprasegmental features, which included stress and intonation. The pronunciation assessments were scored by two trained raters who had expertise in English phonetics and pronunciation instruction. They evaluated recorded performances with a standardized rubric, which primarily focused on segmental accuracy and suprasegmental features. The Learner Confidence Questionnaire, a 20-item survey, was administered before and after the intervention to measure learner confidence, self-efficacy, and willingness to speak in class. The items were reviewed by two English linguistics experts for content validity. The internal consistency reliability was assessed by using Cronbach's alpha (0.80), which showed acceptable reliability coefficients for both pre-test and post-test administration.

The continuous practice sessions and class tasks data were recorded for assessment. Besides this, the class participation, oral tasks, and teacher evaluations were used to triangulate outcomes. The qualitative data included teacher observation notes and weekly learner reflection logs. These were analysed by using thematic analysis with the help of a systematic coding procedure for generating codes and relevant themes. The teacher documented classroom engagement activities, frequency of voluntary participation, and observable improvements in learners' output. There were also weekly reflection logs, which were collected from the experimental group. It meant to capture learner experiences, challenges, and perceived benefits of using AI tools.

The intervention was embedded within the eight-week Speaking and Pronunciation course at the Language Center. In the orientation Week 1, both groups completed their pre-tests and questionnaires. The experimental group received a demonstration of ELSA Speak and Speechling and guidance on daily usage (minimum 15 minutes of practice). In the integration phase from weeks (2-7), the students of the experimental group were given practice sessions both in class (for targeted pronunciation exercises) and at home (for individualized practice).

Teachers monitored engagement through dashboards. The students of the control group were taught the syllabus and practiced pronunciation with drills and speaking aloud activities. In Week 8, assessments were included, a post-test, and questionnaires were part of the assessment phase. Learner reflections and teacher observation notes were collected for analysis. To collect data from teachers and learners for this research study, permission was obtained from the director of the language center. Participation was voluntary, and informed consent was also obtained from the learners. The data were anonymized, and the participants were informed about the objective of the study. The learners were also informed that their course grades would not be affected by their involvement in the study.

Results and Findings

In terms of experimental pronunciation accuracy, AI tools demonstrated greater improvement in pronunciation scores compared to the control group. Independent samples t-tests confirmed significant differences (P < .001). Independent samples *t*-tests were appropriate for comparing means between two separate, unrelated groups, whereas paired-samples *t*-tests were required when the same participants were measured at two time points (e.g., pre-test/post-test) or when observations were otherwise matched.

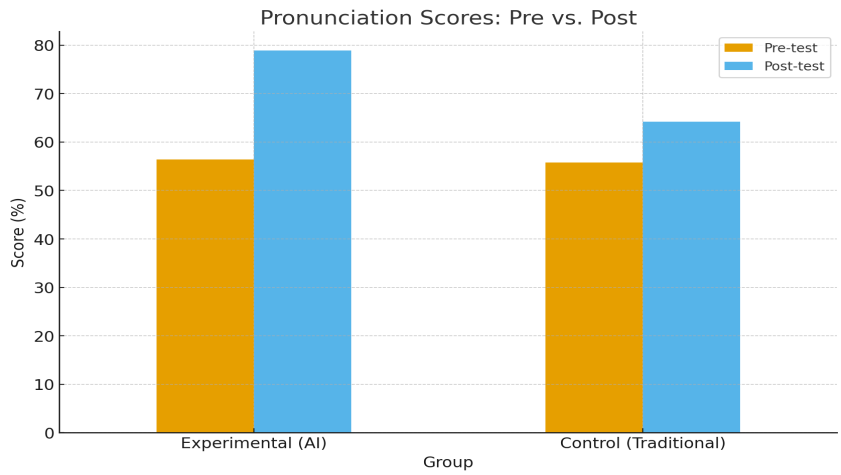

Figure 1. Pre-test and post-test pronunciation scores of the experimental and control groups.

Figure 1 displays pre-test vs. post-test performance. For both groups, the experimental AI group moved from an average proficiency level to substantially higher post-test scores, while the control group showed more modest gains. The independent samples t-test on post-test scores was statistically significant (P < .001), indicating that learners who used ELSA Speak improved their pronunciation accuracy more than their peers, who relied solely on traditional practice and

interpretation. The pattern is consistent with the study's framework. AI tools likely increased form-focused exposure and the quality of feedback events; both segmental and suprasegmental features were also enhanced.

At first, the analysis concerned the heterogeneity of treatment effects with the help of AI tools like ELSA Speak and Speechling. The differences in responses were based on the anxiety levels, prior exposure to computer-mediated tools, or participation patterns for pronunciation practice. The results of the pre-test were used to make profiles of students who needed the most intervention with lower proficiency or higher anxiety levels. It was observed that learners in the experimental group demonstrated greater improvements after learning with AI tools for pronunciation practice. The practice time was a plausible predictor of improvement and linguistic gains. The practice sessions differentiated active and constructive behaviours (e.g., with guided retries, use of exemplars, self-explanations, and also shadowing with self-monitoring from merely passive listening. It was noticed that intense interactive sessions with feedback enhanced cognitive uptake by the learners.

An emphasis on the learning dynamics added explanatory depth, which was beyond pre-post contrasts. The pedagogical counter measures were employed to track progress from segmental targets to suprasegmental patterns, alternating participation cycles to refresh attention, or reconfiguring tasks to vary lexical and prosodic difficulty. These instructional adjustments were consciously made to sustain progress.

The learners 'confidence survey indicated significant gains in the experimental group. Self-reported confidence increased from 2.7 to 4.1in the experimental group as compared to 2.8 to 3.3 in the control group.

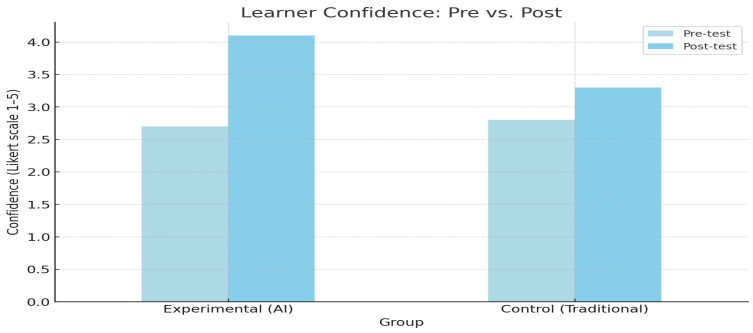

Figure 2: Pre-test and post-test confidence scores of experimental and control groups

As shown in Figure 2. Self-reported confidence rose from 2.7 to 4.1 points in the experimental group to 2.83 points in the control group (scale: 1-5). A paired

samples t-test for the experimental group was significant (P < .001), indicating meaningful gains in perceived ability and willingness to speak. The confidence gains, aligned with the affective filter account, suggested that private, non-judgmental practice appears to reduce anxiety and support willingness to communicate. Usage analytics indicated a strong positive correlation between daily practice time and pronunciation gains (r = 0.62, P < 0.01).

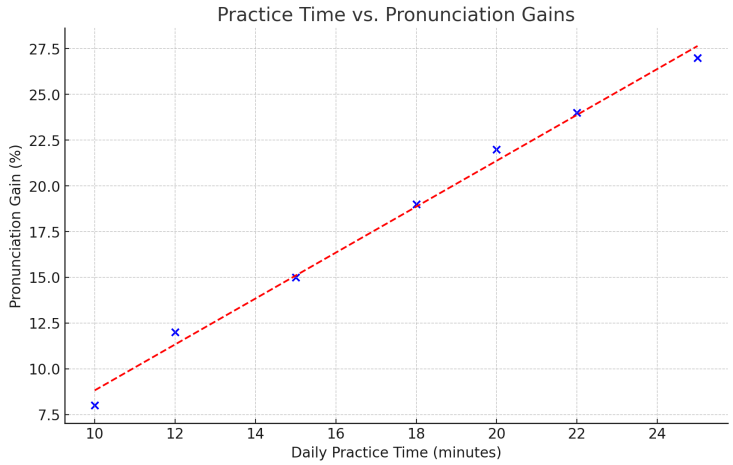

Figure 3. Correlation between Daily Practice Time and Pronunciation Gains

Results and analysis

Figure 3 illustrates A moderate to strong positive correlation between daily practice time and pronunciation gains (r = .62, P< .01). In practical terms, usage explained 38% of the variance in pronunciation gains associated with differences in daily practice time ($r^2 \approx .38$). This finding suggested that learners who engaged in more frequent daily practice demonstrated substantially greater improvement in pronunciation outcomes.

This pattern supported the engagement mechanism proposed in the framework. Qualitative triangulation, teacher observations, and learner reflections converged on three themes: learners valued immediate private feedback and the ability to retry without social pressure, echoing the affective filter account, motivation via micro practice and gamification, streaks of visible progress, and short, achievable tasks supported persistence and self-regulated learning (SRL). SRL behaviors, including planning, monitoring, reflection, free implementation, frictions, intermittent connectivity, and occasional feedback perception mismatches (i.e., where learners doubt the app's outcome), highlighted the need for teacher planning, adjusting tasks, and modelling how to respond to feedback. These themes substantiated the quantitative outcomes and clarified the boundary conditions under which AI tools were most effective. Taken together, these findings

suggested that AI-based speech recognition held considerable promise and potential to improve pronunciation accuracy and enhance learner confidence and satisfaction, especially when paired with careful pedagogical design.

Discussion

The results indicated that AI-assisted pronunciation training improved pronunciation accuracy. It further supports the idea that computer-assisted pronunciation tools enhance phonological awareness, immediate, individualized feedback in contrast to traditional instructions (P<.001). It highlighted that the consistent, form-focused practice with the ELSA Speak facilitated the accurate speech patterns of the learners. It also helped them identify and articulate correct pronunciation errors. Furthermore, the experimental group experienced a marked boost in confidence ratings; this showed that the non-judgmental, low-stakes atmosphere afforded by AI-driven practice likely reduced anxiety and gave way to a greater inclination to communicate with confidence. The implications of regular practice and subsequent performance stress showed the importance of learners' involvement in the technology-enhanced language learning environments. It provided a framework for continuous participation with the help of features similar to tracking progress and small learning tasks. Importantly, the findings also implied a counterbalance of instruction to learners' independence. In the absence of such guidance, learner engagement would likely have remained superficial, with an increased risk of misinterpreting or inadequately responding to system-generated feedback.

The importance of competence in both individual sounds (segmental elements) and stress, rhythmic intonation (suprasegmental features) allowed differentiation between phoneme accuracy and prosodic control. The diversified intonation enhanced segmental intelligibility in pronunciation activities. Differentiation at the level of individual sounds

was significantly enhanced through minimal pair training. The suprasegmental features were heightened through shadowing techniques, rhythmic group, and modulation of vocal pitch, which highlighted the greater impact of the AI feedback. It also demonstrated the mechanism by which the students' speaking and pronunciation deficiencies were addressed. A great improvement in boosting self-confidence and reducing hesitation was accomplished through the ASR-administered activities, such as presentations, dialogues, and spontaneous responses. The practice of stress and intonation under a controlled mental effort established a link with microscopic training to macroscopic comprehension, which had been interpreted as a helpful indicator in achieving communicative competence. Observation of the evolution of major pronunciation errors from regular to irregular occurrence demonstrated proper pronunciation development

that represented internal sound system restoration rather than mere task-specific practice. Hence, both the individual sound features (segmental) and the prosodic features, or suprasegmental features, identified aspects of pronunciation that reflected the utmost response to AI-assisted training with an objective related to communication clarity.

Overall comprehensibility emerged as a marked outcome of the intervention, which indicated students' enhanced macro-level understanding at the outset of the analysis. It supplied visual, auditory, and textual content that catered to diverse learning preferences, which was in line with the multimodal learning theory for cognitive efficiency and knowledge retention (Giannakos & Cukurova, 2023). The data showed that when AI pronunciation exercises were combined with traditional classroom teaching, students produced more accurate stress and pitch variations. Such blending suggested that the impact of AI resources was most effective when they worked alongside a formal language education system. The findings supported a hybrid institutional framework in which AI-driven instruments were supplementary to teacher-led education, by offering personalized, evidence-based observation that advanced language skill improvement and successful learning progress.

Conclusion

The results demonstrated that AI–based tools (ELSA Speak, Speechling) have the potential to improve the pronunciation accuracy and build learners` confidence as compared to conventional methods of teaching pronunciation in language classes. It also helped achieve learners' autonomy, increased motivation, and engagement, and reduced speaking hesitation. The use of innovative pedagogical strategies included optional add-ons for maximizing practice sessions. Future research should incorporate delayed post-tests and adopt multi-site, longitudinal designs to examine the durability of learning gains over time, and enhance the generalisability of findings across diverse institutional contexts, learner profiles, and instructional settings. In conclusion, authentic verbal communication and language development can be embraced with teachers implementing a more harmonious learning environment by integrating the technological tools with established teaching practices.

References

Abdelhalim, S. M., & Alsehibany, R. A. (2025). Integrating AI-powered tools in EFL pronunciation instruction: Effects on accuracy and L2 motivation. *Computer Assisted Language Learning*, 1-25.

Abimanto, D., & Sumarsono, W. (2024). Improving English pronunciation with AI speech-recognition technology. *Acitya: Journal of Teaching and Education*, *6*(1), 146-156.

AbuSahyon, A. S. A. E., Alzyoud, A., Alshorman, O., & Al-Absi, B. (2023). AI-driven technology and chatbots as tools for enhancing English language learning in the context of second language acquisition: A review study. *International Journal of Membrane Science and Technology*, *10*(1), 1209-1223.

Ahadi Kalashi, S. (2023). Review of the noticing hypothesis from two cognitive and ecological perspectives. *Journal of New Advances in English Language Teaching and Applied Linguistics*, *5*(1), 1207-1218.

Amrate, M., & Tsai, P. H. (2025). Computer-assisted pronunciation training: A systematic review. *ReCALL*, *37*(1), 22-42.

Aryanti, R. D., & Santosa, M. H. (2024). A systematic review of artificial intelligence applications for enhancing EFL students' pronunciation skills. *The Art of Teaching English as a Foreign Language (TATEFL)*, *5*(1), 102-113.

Azam, M. K. S. (2011). Speech generation by artificial intelligence systems: Issues and challenges. *NUML Journal of Critical Inquiry*, *9*(1), 62-68.

Caiza, G., Villafuerte, C. & Guanuche, A. (2025). Interactive application with virtual reality and artificial intelligence for improving pronunciation in English learning. *Applied Sciences*, *15*(17), 9270.

Culduz, M. (2023). Gamification in English language teaching (ELT): A comprehensive review of theory and practice. *International Journal of Social and Humanities Sciences Research (JSHSR)*, *10*(100), 2663-2675.

Dai, Y., & Wu, Z. (2023). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning*, *36*(5-6), 861-884.

Dennis, N. K. (2024). Using AI-Powered speech recognition technology to improve English pronunciation and speaking skills. *IAFOR Journal of Education*, *12*(2), 107-126.

Dja'far, V. H., & Hamidah, F. N. (2024). Improving English pronunciation skills through AI-based speech recognition technology. *Ethical Lingua: Journal of Language Teaching and Literature*, *11*(2).

Fountoulakis, M. S. (2024). Evaluating the impact of AI tools on language proficiency and intercultural communication in second language education. *International Journal of Second and Foreign Language Education*, *3*(1), 12-26.

Giannakos, M., & Cukurova, M. (2023). The role of learning theory in multimodal learning analytics. *British Journal of Educational Technology*, *54*(5), 1246-1267.

Gusrianto, E. (2025). Exploring learning experience with ELSA speak for independent learning: A case study. In *International Conference on Advances in Humanities, Education and Language,* 219-229. Atlantis Press.

Hao, T., Wang, Z., & Ardasheva, Y. (2021). Technology-assisted vocabulary learning for EFL learners: A meta-analysis. *Journal of Research on Educational Effectiveness*, *14*(3), 645-667.

Karim, S. A., Hamzah, A. Q. S., Anjani, N. M., Prianti, J., & Sihole, I. G. (2023). Promoting EFL students' speaking performance through ELSA speak: An artificial intelligence in English Language Learning. *Journal of Languages and Language Teaching*, *11*(4), 655-668.

Kawaguchi, S. (2021). Second language acquisition and digital learning in Asia. *Asiatic: IIUM Journal of English Language and Literature*, *15*(1), 36-52.

Lantolf, J. P., & Xi, J. (2023). Digital language learning: A sociocultural theory perspective. *TESOL Quarterly*, *57*(2), 702-715.

Li, P., & Lan, Y. J. (2022). Digital language learning (DLL): Insights from behavior, cognition, and the brain. *Bilingualism: Language and Cognition*, *25*(3), 361-378.

Lv, X., Ren, W., & Xie, Y. (2021). The effects of online feedback on ESL/EFL writing: A meta- analysis. *The Asia-Pacific Education Researcher*, *30*(6), 643-653.

Ma, Q., Mei, F., & Qian, B. (2024). Exploring EFL students' pronunciation learning supported by corpus-based language pedagogy. *Computer Assisted Language Learning*,21(2), 1–27.

Mahmood, R. Q., & Bui, H. P. (2024). EFL learners' perceptions of pronunciation corrective feedback: Insights from synchronous high-variability phonetic training. In *Innovations in technologies for language teaching and learning*, 95–119. Springer Nature Switzerland.

Mohammad Karimi, E. (2024). Exploring the use of artificial intelligence in promoting English language pronunciation skills. *LLT Journal: A Journal on Language and Language Teaching*, *27*(1), 98-115.

Mora, J. C., & Mora-Plaza, I. (2023). From research in the lab to pedagogical practices in the EFL classroom: The case of task-based pronunciation teaching. *Education Sciences*, *13*(10), 1042.

Mubarok, M. Z., & Aziez, F. (2024). AI and non-AI tools in teaching English pronunciation to EFL learners. *Canadian Journal of Language and Literature Studies*, *4*(5), 22-33.

Nazir, S., Javed, F., & Baig, S. (2025). Integrating AI-Powered tools for pronunciation improvement: A study of ESL Learners' perceptions and learning experiences. *Journal of Development and Social Sciences*, *6*(3), 757-765.

Ngo, T. T. N., Chen, H. H. J., & Lai, K. K. W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, *36*(1), 4-21.

Pham, V. T. T., & Pham, A. T. (2025). English major students' satisfaction with ELSA speak in English pronunciation courses. *PloS one*, *20*(1), 23-29.

Pennington, M. C., & Rogerson-Revell, P. (2019). English pronunciation teaching and research. *Londres: Palgrave Macmillan*, *10*, 978-988.

Ping, L., & Tao, N. (2025). Innovative approaches to English pronunciation instruction in ESL environments: Integration of multi-sensor detection and advanced algorithmic feedback. *Interactive Learning Environments*, 18(2),1-21.

Rahayuningsih, R., Hartono, R., & Wahyuni, S. (2025). Leveraging ELSA speak for enhanced speaking proficiency: AI-education (AIEd) tool at private university. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, *13*(1), 1073-1088.

Sari, N. (2023). The role of artificial intelligence (AI) in developing English language learners' communication skills. *Journal on Education*, *6*(1), 750-757.

Seddik, M. E. (2025). The impact of AI-Powered language learning tools on second language acquisition: A mixed-methods study. *International Journal of Linguistics, Literature and Translation*, *8*(3), 269-278.

Syed, N. A. (2015). The role of the teacher in second language learning. *NUML Journal of Critical Inquiry*, *13*(1), 71.

Tran, D. K., & Vu, T. K. C. (2024). Investigating learners' perspectives on ELSA speak integration to enhance autonomy and oral language proficiency in English classes. In *Proceedings of the AsiaCALL International Conference*, *6*(2), 182-192.

Vančová, H. (2023). AI and AI-powered tools for pronunciation training. *Journal of Language and Cultural Education*, *11*(3), 12-24.

Zou, B., Liviero, S., Wei, K., Sun, L., Qi, Y., Yang, X., & Fu, J. (2021). Case study 11, mainland China: The impact of pronunciation and accents in artificial intelligence speech evaluation systems. In *Language Learning with*

Technology: Perspectives from Asia, 223-235. Singapore: Springer Nature Singapore.