RESEARCH ARTICLE

JENER Journal of Empirical and Non-Empirical Research
Volume 2, Issue 10, Page 92-106 | Article No. 10
DOI: 10.4898/jenerv2i10.a10
ISSN: 7895-9216
Journal homepage: https://jenerjournal.com
Journal email: editor@jenerjournal.com

JENER Journal
JENER Journal of Empirical and Non-Empirical Research

AI Versus Human Raters: Effectiveness of Automated Feedback in L2 Pronunciation Assessment

Prof. Langat Christophers Andrew[1]

[1]Commissioner (Kenya Commission for University Education)

ARTICLE INFORMATION

Article history:
Published: October 2026

Keywords:
Artificial Intelligence (AI)
Automatic Speech Recognition (ASR)
L2 Pronunciation Assessment
Automated Feedback
Human Rater Feedback
Pronunciation Accuracy
Suprasegmental Features
Learner Perceptions

ABSTRACT

This study investigated the comparative effectiveness of artificial intelligence (AI)-based feedback and human rater feedback in second language (L2) pronunciation assessment. Pronunciation plays a critical role in intelligibility and communicative competence, yet traditional assessment methods relying on human evaluators have been constrained by issues of time, cost, and inter-rater variability. Recent advances in automatic speech recognition (ASR) and machine learning have introduced AI-driven tools capable of delivering immediate, consistent, and scalable pronunciation feedback; however, questions remained regarding the pedagogical quality and effectiveness of such automated systems in comparison to human judgment. The study evaluated the accuracy, impact, and perceived usefulness of AI-generated feedback relative to human-provided feedback among university-level English as a foreign language (EFL) learners. Specifically, it examined differences in learners' pronunciation improvement, analyzed the alignment between AI and human ratings, and explored learners' perceptions of both feedback types. By focusing on both segmental and suprasegmental features, the research provided a comprehensive assessment of pronunciation development. The findings clarified that AI-based feedback can serve as a valuable complementary tool to human raters, particularly in providing immediate and consistent input, although it was less effective in addressing complex prosodic and context-dependent features. The results contributed to applied linguistics and language pedagogy by offering empirical evidence to guide the integration of AI technologies into pronunciation instruction, supporting more effective, scalable, and learner-centered assessment practices.

1. Introduction

Pronunciation is a fundamental component of second language (L2) learning, as it directly affects intelligibility, comprehensibility, and effective oral communication. In applied linguistics, pronunciation assessment has traditionally relied on human raters, such as teachers or trained evaluators, who provide feedback based on their phonetic knowledge and pedagogical judgment. While human-delivered feedback is often considered rich and context-sensitive, it can also be time-consuming, costly, and subject to inter-rater variability and inconsistency (Amrate & Tsai, 2024).

Recent advancements in artificial intelligence (AI), particularly in automatic speech recognition (ASR) and machine learning, have led to the development of automated pronunciation assessment systems. These systems are increasingly used in computer-assisted pronunciation training (CAPT) and language learning applications to analyze learners' speech and provide immediate feedback on segmental and suprasegmental features. Research indicates that AI-based pronunciation tools can offer scalable, consistent, and timely feedback, which is especially valuable in large or autonomous learning contexts (El Kheir et al., 2023).

Empirical evidence suggests that ASR-mediated feedback has a positive impact on L2 pronunciation development. A recent meta-analysis by Ngo et al. (2024) found that automated pronunciation feedback produced a moderate effect on learners' pronunciation improvement, particularly when the feedback was explicit and form-focused. However, the authors also noted that many automated systems still struggle to capture prosodic features and contextual nuances that human raters can more readily identify.

Despite the growing use of AI-based assessment tools, questions remain regarding the effectiveness and pedagogical quality of automated feedback compared to human rater feedback. Learner perception studies suggest that while students appreciate the immediacy and objectivity of AI feedback, they often value human feedback for its explanatory depth and personalized guidance (Susanto, 2025). Moreover, recent research comparing automated scores with human ratings shows promising alignment in certain pronunciation aspects, though discrepancies persist in diagnosing complex or discourse-level pronunciation issues (Parikh et al., 2026).

Given these developments, further research is needed to systematically compare AI-based and human rater feedback in L2 pronunciation assessment. Understanding their relative effectiveness can inform more balanced and pedagogically sound integration of automated tools into pronunciation instruction and assessment practices.

1.2 Statement of the Problem

The rapid advancement of artificial intelligence (AI) technologies has led to their increasing adoption in second language (L2) pronunciation assessment and instruction. AI-based pronunciation tools, often powered by automatic speech recognition (ASR), are now widely used in language learning applications and computer-assisted pronunciation training environments. These tools are

promoted for their ability to provide immediate, consistent, and scalable feedback to learners. However, despite their growing popularity, there is still limited empirical research that systematically compares AI-generated pronunciation feedback with feedback provided by human raters, who have traditionally played a central role in pronunciation assessment. This lack of direct comparison raises concerns about the evidence base supporting the widespread implementation of AI tools in pronunciation pedagogy.

Furthermore, there is ongoing uncertainty regarding the pedagogical quality, accuracy, and instructional usefulness of AI-generated feedback. While AI systems can efficiently detect certain pronunciation errors, particularly at the segmental level, questions remain about their ability to provide meaningful, explanatory, and context-sensitive feedback that supports learner development. Human raters, such as teachers or trained evaluators, are often able to tailor feedback to individual learner needs and explain errors in ways that facilitate noticing and improvement. The extent to which AI feedback can match this level of pedagogical effectiveness remains unclear, especially for more complex pronunciation features such as stress, rhythm, and intonation.

As a result of these unresolved issues, there is a lack of consensus in the field of applied linguistics on whether AI-based pronunciation feedback can effectively replace or should merely supplement human feedback in L2 pronunciation assessment. Some researchers argue that AI tools have the potential to reduce teacher workload and increase learner autonomy, while others caution against overreliance on automated systems that may oversimplify pronunciation assessment. This absence of agreement highlights the need for further empirical investigation to determine the relative effectiveness of AI and human feedback, as well as how both can be optimally integrated to support L2 pronunciation learning.

1.3 Purpose of the Study

The purpose of this study is to examine and compare the effectiveness of AI-based pronunciation feedback and human rater feedback in second language (L2) pronunciation assessment. Specifically, the study seeks to determine how well automated pronunciation feedback supports learners' pronunciation development in comparison with feedback provided by human raters, such as language teachers or trained evaluators. By analyzing learners' pronunciation performance following exposure to both types of feedback, the study aims to identify similarities and differences in their impact on pronunciation accuracy and overall spoken intelligibility.

In addition, the study aims to evaluate the pedagogical usefulness of AI-generated feedback by considering its clarity, accuracy, and relevance from the learners' perspective. Understanding whether AI feedback can facilitate noticing of pronunciation errors and promote meaningful improvement is essential, particularly as such tools are increasingly integrated into language learning contexts. Through this comparative approach, the study also seeks to clarify whether AI-based pronunciation assessment can serve as a reliable alternative to human feedback or function more effectively as a complementary tool. Ultimately, the findings are expected to contribute empirical evidence that informs pronunciation assessment practices and supports informed decision-making regarding the integration of AI technologies in L2 instruction.

1.4 Research Objectives

The study aimed to compare the effectiveness of AI-based pronunciation feedback and human rater feedback in L2 pronunciation assessment. Specifically, it sought to compare the accuracy of pronunciation feedback provided by AI-based systems and human raters. The study also examined learners' pronunciation improvement after receiving AI-based feedback in comparison to human rater feedback. In addition, it investigated learners' perceptions of AI-generated feedback and human-provided feedback in order to understand their perceived usefulness and acceptability. Finally, the study identified the strengths and limitations of AI-based pronunciation assessment based on performance outcomes and learner responses.

1.5 Research Questions

This study addressed the following research questions:

- How effective was AI-based feedback in L2 pronunciation assessment when compared to human rater feedback?
- Was there a significant difference in learners' pronunciation improvement after receiving AI-based feedback and human rater feedback?
- How did L2 learners perceive AI-based pronunciation feedback in comparison to feedback provided by human raters?

1.6 Significance of the Study

This study is significant for L2 learners as it highlights the potential of AI-based pronunciation assessment to provide timely, consistent, and accessible feedback that can support independent practice and continuous pronunciation improvement. By comparing AI-generated feedback with human rater feedback, the study offers insights into how learners can benefit from immediate corrective input, especially in contexts where teacher feedback may be limited due to time or class size.

For language teachers, the findings of this study provide practical support for instructional and assessment practices by clarifying the strengths and limitations of AI-based pronunciation tools. Understanding how AI feedback compares with human feedback can help teachers make informed decisions about integrating technology into pronunciation instruction, potentially reducing workload while maintaining feedback quality.

The study is also significant for curriculum designers, as it offers empirical evidence to guide the effective and pedagogically sound integration of AI tools into language curricula. By identifying areas where AI-based pronunciation assessment is most effective, curriculum developers can design blended approaches that combine technological tools with human instruction to enhance learning outcomes.

Finally, this study contributes to researchers in the fields of ICT and applied linguistics by extending existing knowledge on automated language assessment and feedback. The findings provide empirical data that can inform future research on AI-assisted language learning and support the development of more effective and learner-centered pronunciation assessment technologies.

1.7 Scope and Delimitations

This study focused on L2 learners, specifically university students learning English as a foreign language (EFL), to examine the impact of AI-based and human rater feedback on pronunciation improvement. The research concentrated on selected pronunciation

features, including both segmental aspects (individual sounds) and suprasegmental aspects (stress, intonation, and rhythm), to provide a comprehensive assessment of learners' oral performance. The study utilized specific AI pronunciation assessment tools chosen for their accessibility, reliability, and instructional relevance, rather than all available technologies, which limits the generalizability of the findings to these tools. Additionally, the study was conducted over a limited intervention period, which may have influenced the extent of observable pronunciation improvement and the long-term applicability of the results. Despite these limitations, the scope allowed for a focused and manageable investigation of the comparative effectiveness of AI and human feedback in pronunciation assessment.

1.8 Operational Definition of Terms

To ensure clarity and consistency throughout this study, key terms are defined operationally as follows:

Artificial Intelligence (AI): In this study, artificial intelligence refers to computer systems or software that are capable of performing tasks that normally require human intelligence. Specifically, AI is used to analyze and evaluate spoken language, identify pronunciation errors, and generate feedback for learners in a way that simulates human judgment. Examples include machine learning algorithms, automatic speech recognition (ASR) systems, and speech analysis tools integrated into language learning applications.

Automated Pronunciation Assessment: This term refers to the use of AI-based systems to evaluate learners' pronunciation without direct human intervention. Automated pronunciation assessment involves analyzing spoken input to detect segmental errors (individual sounds), suprasegmental errors (stress, rhythm, intonation), and overall speech intelligibility, and then providing feedback or scores that guide learners toward improvement. In this study, it specifically refers to the selected AI tools employed to give immediate and structured pronunciation feedback to participants.

Human Raters: Human raters are qualified individuals, such as language teachers or trained evaluators, who assess learners' spoken performance. They use established pronunciation rubrics and their linguistic expertise to identify errors, provide corrective feedback, and evaluate overall intelligibility. Human raters serve as a benchmark in this study for comparing the effectiveness and quality of AI-generated pronunciation feedback.

Pronunciation Feedback: Pronunciation feedback refers to information provided to learners regarding the accuracy, clarity, and correctness of their spoken language. This feedback can be corrective, descriptive, or evaluative, and may include guidance on segmental articulation, stress patterns, intonation, and rhythm. In this study, pronunciation feedback encompasses both the responses generated by AI systems and those provided by human raters, which are compared for effectiveness.

L2 Pronunciation: L2 pronunciation refers to the way learners produce speech sounds, stress, rhythm, and intonation in a language that is not their first language (second language). This includes both segmental elements, such as vowels and consonants, and suprasegmental elements, such as sentence stress and prosody. The study focuses on L2 English pronunciation among EFL university students as the target area for assessment and feedback.

2. Review of Related Literature

2.1 Theoretical Framework

Second Language Acquisition (SLA) theories provide the foundation for understanding how learners acquire a second language and how feedback facilitates learning, particularly in the development of L2 pronunciation (Schmidt, 1990; Gass & Mackey, 2020). SLA research emphasizes the interplay between cognitive, social, and environmental factors in language acquisition. This includes how learners notice errors, process corrective feedback, and integrate accurate forms into their interlanguage system. These theoretical principles are critical for examining the comparative effectiveness of AI-based automated feedback versus traditional human rater feedback in pronunciation assessment, as they frame both the mechanisms of learning and the expected outcomes.

The Noticing Hypothesis, proposed by Schmidt (1990), asserts that learners must consciously notice gaps between their current language production and the target language to acquire new forms. In pronunciation learning, this conscious awareness is essential because learners often mispronounce sounds without realizing their deviation from target norms. Noticing is the cognitive trigger that enables learners to recognize their errors and take corrective action. Within this framework, feedback whether provided by AI or human raters serves as a critical mechanism to draw attention to these discrepancies and support learning.

AI-based feedback can enhance noticing by providing immediate, explicit signals of pronunciation errors, both at the segmental level (individual sounds) and suprasegmental level (intonation, stress, rhythm). These tools often highlight mispronunciations visually, provide auditory models, or assign accuracy scores, allowing learners to identify specific areas for improvement (Farrús, 2023). While AI feedback offers immediacy and consistency, human raters contribute nuanced explanations contextualized to the learner's linguistic background and speech patterns. For example, teachers can clarify why a specific vowel or consonant is misproduced and offer strategies to correct it, which may facilitate deeper cognitive processing (Ngo, Chen, & Lai, 2024).

The Interactionist Perspective emphasizes that language acquisition occurs most effectively through meaningful communicative interaction that includes comprehensible input, opportunities for output, and responsive feedback (Long, 1996). In pronunciation learning, interaction allows learners to adjust their speech in real-time through negotiation of meaning, clarification requests, and recasts. Human raters can provide tailored, context-dependent feedback during communicative activities, which supports learners in refining their pronunciation. Interactionist theory highlights the social dimension of language learning, suggesting that feedback is not merely corrective but also communicative, promoting learner engagement.

AI feedback, while lacking true social interaction, can simulate aspects of interaction by providing immediate corrective signals that encourage learners to self-correct and repeat speech. Some AI tools provide adaptive feedback based on learner performance, allowing for iterative practice that mimics interaction to a limited degree (Sun, 2023). However, AI cannot replicate the social and pragmatic nuances of human-mediated feedback, such as empathetic reinforcement, modeling discourse-level intonation, or responding to learner affect and motivation. This limitation underscores the importance of considering AI as a complementary rather than a complete replacement for human feedback in pronunciation instruction.

Corrective feedback theory in SLA explains the mechanisms through which learners detect and modify errors in interlanguage production. Corrective feedback can be explicit (direct correction), implicit (recasts or prompts), or metalinguistic (providing explanations of errors), and it facilitates learners' awareness of phonetic and prosodic deviations (Lyster & Saito, 2010). Human raters have the advantage of delivering multi-layered feedback that adapts to learners' proficiency, enabling strategic scaffolding of complex pronunciation features. AI feedback, on the other hand, is often limited to error detection and scoring but can provide consistency and immediacy that human raters may not match.

Empirical studies show that both AI and human feedback can improve pronunciation, especially when learners actively engage with the corrections. For instance, ASR-based feedback systems have demonstrated measurable gains in segmental pronunciation accuracy, whereas human feedback often excels in addressing suprasegmental aspects such as stress, rhythm, and intonation (Ngo et al., 2024). This distinction suggests that human and AI feedback have complementary strengths: AI offers efficient, frequent, and objective assessment, while human raters provide interpretive, context-sensitive guidance.

The Technology Acceptance Model (TAM) complements these SLA perspectives by addressing learner engagement with AI tools. TAM posits that perceived usefulness and perceived ease of use influence whether learners adopt a technology and how effectively they engage with it (Davis, 1989). Studies of AI-based pronunciation assessment show that learners who perceive the system as user-friendly and beneficial are more likely to practice consistently, leading to greater exposure to corrective feedback and improved pronunciation outcomes (Pronunciation Trainer, 2025). TAM thus adds an affective and motivational lens, highlighting how technology design and learner perception influence learning efficacy.

Integrating these theoretical perspectives provides a multidimensional framework that incorporates cognitive, social, and technological factors in pronunciation learning. SLA theories (Noticing Hypothesis and Interactionist Perspective) explain why and how feedback promotes acquisition, corrective feedback theory specifies the mechanisms of error detection and remediation, and TAM accounts for learner engagement and the adoption of AI tools. Together, these frameworks justify the comparative evaluation of AI and human feedback and inform the design of interventions aimed at optimizing pronunciation development.

This framework also emphasizes that the effectiveness of AI-based pronunciation assessment should be evaluated beyond mere accuracy. Factors such as the timing of feedback, interpretability, learner engagement, and alignment with instructional goals are critical for understanding its pedagogical value (Farrús, 2023; Ngo et al., 2024). Human feedback, while potentially less scalable, contributes unique value by addressing learner-specific challenges and providing nuanced corrective strategies.

Furthermore, the integration of SLA and TAM highlights the importance of learner perception and motivation in feedback-mediated learning. Even highly accurate AI feedback may have limited impact if learners do not perceive it as relevant or actionable. Conversely, learners who actively engage with AI feedback and view it as helpful may achieve improvements comparable to those obtained through human feedback. This underscores the need to study both objective performance outcomes and subjective learner responses in pronunciation research.

The theoretical framework, therefore, positions this study to explore the interplay between feedback type, learner engagement, and pronunciation outcomes. By considering cognitive, social, and technological dimensions simultaneously, it allows for a comprehensive understanding of how AI-based and human feedback function individually and interactively in supporting L2 pronunciation learning. Finally, by grounding the study in these theories, researchers can not only compare the effectiveness of AI and human feedback but also generate insights into how best to integrate AI tools into existing pedagogical practices. The framework informs the design of pronunciation tasks, selection of assessment metrics, and evaluation of learner perceptions, ensuring that the study addresses both theoretical and practical considerations in applied linguistics.

2.2 Pronunciation in Second Language Learning

Pronunciation is a central component of second language (L2) learning, as it directly affects intelligibility and comprehensibility—the ability of listeners to understand spoken messages accurately and efficiently (Derwing & Munro, 2015). Intelligibility refers to how easily a listener can identify intended words and sounds, while comprehensibility encompasses the overall clarity of speech, including rhythm, stress, and prosody (Munro & Derwing, 2018). Research shows that even when learners possess adequate vocabulary and grammar, poor pronunciation can significantly hinder communication, resulting in misunderstandings or communicative breakdowns. Therefore, pronunciation is not merely a decorative aspect of speech but a functional skill essential for effective communication in L2 contexts.

Pronunciation in L2 learning encompasses segmental and suprasegmental features. Segmental features include individual speech sounds, such as vowels and consonants, and the correct articulation of phonemes (Celce-Murcia, Brinton, & Goodwin, 2010). Mastery of segmental features ensures that individual words are recognizable, directly contributing to intelligibility. Many learners struggle with producing L2 phonemes that do not exist in their first language (L1), which can lead to persistent mispronunciations and reduced understanding by native speakers. Segmental training is therefore a foundational aspect of pronunciation instruction, particularly in the early stages of L2 learning.

Suprasegmental features, by contrast, involve prosodic elements such as stress, intonation, rhythm, and pitch patterns that occur across syllables, words, and sentences (Celce-Murcia et al., 2010; Derwing & Munro, 2015). Suprasegmental features are crucial for conveying meaning, emotion, and discourse-level distinctions. For example, incorrect word stress can cause listeners to misidentify a word entirely, while inappropriate intonation can lead to confusion regarding sentence type (question vs. statement) or speaker attitude. Studies indicate that suprasegmental mastery is often more important than perfect segmental production for overall comprehensibility, especially in natural, connected speech (Munro & Derwing, 2018).

Research also highlights the interrelationship between segmental and suprasegmental features. While segmental errors may impede individual word recognition, suprasegmental errors often affect sentence-level intelligibility and listener comprehension. Therefore, comprehensive pronunciation instruction should integrate both aspects rather than focusing exclusively on phoneme articulation. Automated tools for pronunciation assessment increasingly attempt to address both segmental and suprasegmental features;

however, many AI systems remain more accurate for segmental detection than for prosodic evaluation, highlighting the continuing role of human feedback in L2 pronunciation instruction (Farrús, 2023; Ngo, Chen, & Lai, 2024).

Effective pronunciation training also relies on feedback and practice strategies that target both levels. Segmental training may include minimal pair drills, phonetic transcription exercises, and repetition of target phonemes, whereas suprasegmental training may involve stress pattern marking, sentence intonation modeling, and rhythm exercises. Empirical evidence suggests that learners who receive explicit, focused instruction on both features achieve higher levels of intelligibility and comprehensibility than those who practice without targeted feedback (Derwing & Munro, 2015; Celce-Murcia et al., 2010).

Intelligibility and comprehensibility are particularly significant in L2 contexts where learners interact with speakers of varying linguistic backgrounds. For example, research in English as a Foreign Language (EFL) classrooms has shown that learners who produce speech that is intelligible but not native-like in accent can still communicate effectively, underscoring the practical importance of focusing on functional pronunciation rather than perfect imitation of native norms (Derwing & Munro, 2015). This perspective informs the comparative evaluation of AI and human feedback, emphasizing the need for tools and instructional strategies that enhance functional communication.

Pronunciation also affects learners' confidence and willingness to communicate (WTC). Learners who struggle with segmental or suprasegmental features often hesitate to speak, limiting opportunities for practice and interaction. By providing timely corrective feedback, whether from human raters or AI systems, learners can develop greater confidence, ultimately increasing oral fluency and communication success (Munro & Derwing, 2018). Therefore, pronunciation instruction is not solely about accuracy but also about fostering learner autonomy and communicative competence.

The distinction between segmental and suprasegmental features is also critical for the design of assessment tools. Automated pronunciation assessment systems often excel at detecting segmental errors, such as mispronounced vowels or consonants, but are less reliable in evaluating suprasegmental elements like sentence stress and intonation. Human raters, by contrast, can assess both levels holistically, providing corrective guidance that addresses intelligibility and naturalness simultaneously (Farrús, 2023). Understanding these distinctions is vital for evaluating the relative effectiveness of AI and human feedback in L2 pronunciation.

Moreover, the importance of pronunciation is magnified in multilingual and global communication contexts, where comprehensibility often takes precedence over accent reduction. Studies show that learners with intelligible pronunciation, even with an identifiable L1 accent, are generally more successful in communication than those with accurate but inconsistent segmental or suprasegmental production (Derwing & Munro, 2015). This emphasizes that pronunciation instruction and assessment should prioritize communicative effectiveness, guiding both AI and human feedback approaches.

In summary, pronunciation in L2 learning is a multidimensional skill that significantly affects intelligibility, comprehensibility, learner confidence, and overall communicative competence. It encompasses both segmental and suprasegmental features, each contributing uniquely to successful oral communication. Effective instruction and assessment, therefore, require targeted feedback that addresses both aspects, and this theoretical understanding informs the comparative evaluation of AI-based and human feedback systems in L2 pronunciation research.

2.3 Human Rater-Based Pronunciation

Assessment Human rater-based pronunciation assessment remains a central method for evaluating L2 learners' spoken language proficiency, particularly in formal education and research contexts. Human raters, often trained language teachers or linguists, assess learners' pronunciation based on established rubrics or scoring criteria. Their evaluations draw on expertise in phonetics, phonology, and teaching experience to judge both the accuracy and comprehensibility of speech. Rater judgment and expertise are therefore fundamental to ensuring meaningful and pedagogically relevant feedback for L2 learners (Isaacs, 2017).

The expertise of human raters directly influences the quality of pronunciation assessment. Expert raters are able to identify subtle segmental and suprasegmental errors that may escape automated detection systems. They can evaluate pronunciation in context, considering factors such as speech rate, connected speech, and prosodic patterns. Moreover, experienced raters can tailor their feedback to the learner's proficiency level and learning needs, offering metalinguistic explanations or corrective strategies that facilitate deeper understanding (Levis & Grant, 2003). This human interpretive ability remains a significant advantage over AI systems, which may lack sensitivity to context, nuance, and communicative intent.

Despite these advantages, human rater-based assessment faces several reliability and validity challenges. Reliability refers to the consistency of scoring across different occasions or raters, while validity concerns whether the assessment measures what it intends to measure—in this case, intelligibility, comprehensibility, and pronunciation accuracy. Studies have shown that rater judgments can be influenced by subjective factors, such as familiarity with the learner's accent, personal perceptions of "nativeness," or biases related to social or cultural expectations (Munro, Derwing, & Morton, 2006). Consequently, even trained raters may provide inconsistent evaluations unless assessment procedures are carefully standardized and raters are calibrated through training.

Inter-rater variability is a well-documented phenomenon in human rater-based pronunciation assessment. Different raters may assign different scores to the same speech sample, particularly when evaluating suprasegmental features such as stress, rhythm, and intonation, which are inherently more subjective than segmental features (Isaacs & Trofimovich, 2012). Inter-rater variability can be mitigated through detailed scoring rubrics, rater training sessions, and norming procedures that align judgments across raters. Research indicates that when raters receive training in calibration and scoring consistency, the reliability of human pronunciation assessment improves significantly (Levis, 2005).

The validity of human rater judgments is also affected by the constructs being assessed. While raters are generally reliable in evaluating intelligibility, they may differ in how much weight they assign to accent, prosody, or segmental accuracy. For example, some raters may overemphasize native-like pronunciation, whereas others prioritize functional communication. These differences can affect the consistency and interpretability of assessment results, highlighting the need for explicit scoring criteria and clear assessment objectives (Derwing & Munro, 2015).

Despite potential variability, human raters provide qualitative feedback that extends beyond numeric scores. They can explain the nature of errors, suggest corrective strategies, and provide encouragement, all of which are critical for learner motivation and improvement. Such feedback often addresses both segmental and suprasegmental features in a holistic manner, which is essential for communicative competence in L2 contexts (Isaacs, 2017; Munro & Derwing, 2018).

In comparison with automated pronunciation assessment systems, human raters demonstrate flexibility and contextual sensitivity. They can accommodate diverse accents, dialects, and speech styles, recognizing intelligible variations that AI tools might mark as incorrect. Human feedback is also adaptive: raters can adjust their focus depending on the learner's developmental stage or instructional goals, making it a dynamic component of pronunciation instruction (Levis & Grant, 2003; Farrús, 2023).

Human rater-based assessment is particularly valuable when evaluating suprasegmental features, such as sentence stress, intonation, and rhythm. While AI systems are improving in detecting segmental errors, they remain limited in assessing prosody accurately. Human raters, by integrating auditory judgment with linguistic knowledge, can provide nuanced evaluation and guidance that supports learner comprehension and naturalness in speech (Derwing & Munro, 2015; Isaacs & Trofimovich, 2012).

However, the resource-intensive nature of human rater assessment—requiring time, training, and coordination—poses practical constraints. Large-scale or frequent assessment may not be feasible in many educational contexts. These limitations underscore the potential of AI-based tools to complement human assessment by providing consistent, immediate, and scalable feedback, while human raters contribute interpretive depth and contextual accuracy (Ngo, Chen, & Lai, 2024).

In summary, human rater-based pronunciation assessment relies on the expertise and judgment of trained evaluators to provide reliable and valid feedback. While challenges such as inter-rater variability, reliability, and construct validity exist, systematic training, rubric development, and calibration procedures can mitigate these issues. Human raters remain indispensable for evaluating nuanced features of pronunciation, particularly suprasegmental elements, and for providing detailed, pedagogically useful feedback that supports L2 learner development. Their role is crucial in both classroom instruction and research contexts, complementing emerging AI-based assessment technologies.

2.4 AI and Automated Pronunciation Assessment

Artificial intelligence (AI) has increasingly been applied in second language (L2) education, particularly in the area of pronunciation assessment. AI-driven systems offer automated, scalable, and consistent evaluation of spoken language, addressing some of the challenges associated with human rater-based assessment, such as inter-rater variability and resource constraints (Farrús, 2023). By leveraging technologies such as Automatic Speech Recognition (ASR) and machine learning algorithms, AI tools can detect pronunciation errors, provide immediate feedback, and facilitate individualized learning for L2 learners.

Automatic Speech Recognition (ASR) is the core technology underlying many AI pronunciation assessment systems. ASR converts spoken input into text and analyzes it for segmental and suprasegmental accuracy, comparing learner production to target norms (Ngo, Chen, & Lai, 2024). ASR-based systems can identify mispronounced phonemes, syllable stress patterns, and intonation discrepancies. Research demonstrates that ASR technology can provide objective and reproducible assessment, which is particularly valuable for high-volume or remote learning contexts where human raters are unavailable (Sun, 2023). However, while ASR excels at detecting segmental errors, it often struggles with suprasegmental features, such as intonation or sentence rhythm, highlighting the continued need for human oversight in holistic pronunciation evaluation.

Computer-Assisted Pronunciation Training (CAPT) refers to software applications that integrate ASR with interactive learning activities to improve learners' pronunciation. CAPT systems typically provide real-time feedback, visualizations, and practice exercises designed to target specific phonetic features (Saito, 2020). Examples include showing spectrograms of learner speech, displaying pitch contours, or highlighting mispronounced phonemes in color-coded formats. CAPT platforms have been shown to improve both segmental and suprasegmental features of pronunciation, particularly when learners engage in repeated practice and receive feedback that guides self-correction (Farrús, 2023; Sun, 2023). These systems also enhance learner autonomy, allowing practice outside classroom settings.

AI-based pronunciation assessment provides different types of feedback to facilitate learning. One common form is scoring, where the system assigns numerical values to pronunciation accuracy, intelligibility, or fluency. Scores allow learners to monitor their progress over time and identify areas that require further practice (Ngo et al., 2024). Another form is visual feedback, which can include spectrograms, pitch graphs, or articulatory diagrams that show deviations from target production. Visual cues help learners notice errors and understand the nature of their mispronunciations in a more concrete way, aligning with the Noticing Hypothesis (Schmidt, 1990). Corrective prompts are another form of AI feedback, offering immediate suggestions or models for improvement, such as highlighting the correct pronunciation, providing slowed or exaggerated speech models, or prompting the learner to repeat the word or sentence (Saito, 2020; Farrús, 2023).

The immediacy and consistency of AI feedback are among its main advantages. Unlike human raters, who may vary in judgments due to subjectivity, fatigue, or contextual factors, AI systems provide reproducible evaluations and standardized corrective prompts for all learners. This makes AI tools particularly suitable for large classes or distance learning contexts, where individualized human feedback is impractical (Ngo et al., 2024). Moreover, AI feedback supports repeated, self-directed practice, which is critical for developing automaticity in L2 pronunciation and integrating corrective input into learners' interlanguage systems.

However, there are limitations to AI-based assessment. While effective for segmental features, AI systems may not accurately capture suprasegmental patterns such as prosody, intonation, and stress, which are essential for naturalness and overall comprehensibility (Sun, 2023; Derwing & Munro, 2015). Additionally, learners may misinterpret feedback if visualizations or scores are not clearly explained, highlighting the importance of instructional guidance and user training in CAPT environments. The motivational and affective dimensions of feedback such as encouragement and personalized strategies—are also less effectively provided by AI than by human raters.

Recent studies indicate that blended approaches, which combine AI-based feedback with human raters' interpretive guidance, offer the most pedagogically effective solution for pronunciation learning (Farrús, 2023; Ngo et al., 2024). In such approaches, AI systems

provide immediate and objective error detection, while human instructors focus on contextual, prosodic, and strategic guidance, thereby integrating the strengths of both assessment methods. This combination addresses both the cognitive and social dimensions of learning, consistent with SLA theories and feedback frameworks.

In conclusion, AI and automated pronunciation assessment, particularly via ASR and CAPT, provide innovative tools that complement traditional human rater feedback. These systems offer consistent, immediate, and objective evaluation, leveraging scores, visual cues, and corrective prompts to support learner noticing and self-directed practice. While AI excels in segmental assessment and scalability, limitations in prosodic evaluation and motivational support underscore the continued importance of human involvement. Emerging research suggests that a hybrid model integrating AI and human feedback may represent the most effective approach to pronunciation assessment in L2 learning contexts.

2.5 Empirical Studies on AI vs. Human Feedback

The increasing integration of AI tools in second language (L2) pronunciation instruction has prompted a growing body of empirical research comparing the effectiveness of AI-based feedback with traditional human rater feedback. Studies examining these modalities provide insights into their respective strengths, limitations, and learner perceptions, informing pedagogical decisions in applied linguistics and language education.

Studies supporting AI effectiveness demonstrate that AI-based systems, particularly those using Automatic Speech Recognition (ASR) and Computer-Assisted Pronunciation Training (CAPT), can improve segmental pronunciation and contribute to measurable gains in intelligibility. For instance, Ngo, Chen, and Lai (2024) conducted a meta-analysis of ASR-based pronunciation interventions and found that learners who engaged with AI feedback showed significant improvements in vowel and consonant articulation compared to control groups without automated support. Similarly, Sun (2023) reported that EFL learners who used ASR-based CAPT tools achieved higher accuracy in segmental pronunciation after a 6-week intervention, highlighting the potential of AI to provide consistent, immediate, and individualized feedback at scale. These findings suggest that AI systems can serve as effective supplements to classroom instruction, particularly in contexts where human resources are limited.

Other studies emphasize the role of AI in promoting learner autonomy and frequent practice. Farrús (2023) notes that AI feedback allows learners to engage in repeated self-directed pronunciation exercises without relying on constant teacher supervision. This aligns with SLA theories such as the Noticing Hypothesis (Schmidt, 1990), as immediate feedback encourages learners to detect and correct errors independently. The ability of AI tools to provide visual cues, scoring, and corrective prompts also reinforces learner noticing and supports incremental improvement over time.

Despite these advantages, research also highlights limitations of AI feedback, particularly in addressing suprasegmental features such as stress, intonation, and rhythm. While AI systems are highly accurate in detecting segmental errors, their evaluation of prosody remains less reliable due to the complexity of pitch variation, speech rate, and sentence-level prosodic patterns (Derwing & Munro, 2015; Sun, 2023). This limitation reduces AI's ability to fully replicate the nuanced assessments of human raters, who can provide holistic evaluations of both segmental and suprasegmental features and tailor feedback to the learner's context (Isaacs, 2017). Furthermore, AI feedback can lack interpretive depth; learners may receive scores or visual highlights but struggle to understand the underlying nature of their errors or how to correct them effectively (Farrús, 2023).

Several empirical studies also underscore the importance of human oversight when using AI-based feedback. Levis and Grant (2003) argue that human raters are indispensable for providing explanations, metalinguistic guidance, and motivational support, which AI tools cannot replicate. Studies comparing AI and human feedback suggest that learners benefit most when AI systems are used as supplementary tools rather than replacements, combining the scalability and immediacy of AI with the interpretive richness of human assessment (Ngo et al., 2024).

Research on learner perceptions provides additional insights into the effectiveness of AI-based feedback. Studies indicate that learners generally appreciate the immediacy and objectivity of AI feedback but may perceive it as less personalized or less trustworthy than human feedback. For example, Saito (2020) found that learners valued visual feedback and scoring for tracking progress but expressed a preference for human explanations when addressing more complex prosodic or context-dependent errors. This highlights the need for instructional scaffolding alongside AI tools to ensure that learners interpret feedback correctly and remain motivated.

Some studies explore comparative effectiveness in integrated settings, where learners receive both AI and human feedback. Farrús (2023) reports that such blended approaches yield the highest gains in pronunciation accuracy and learner satisfaction. AI provides continuous practice and objective error detection, while human instructors focus on interpretive guidance, prosodic features, and individualized correction strategies. This hybrid model aligns with SLA principles, as it addresses cognitive, social, and affective dimensions of language learning simultaneously.

Critically, while empirical evidence supports AI's capacity to improve segmental pronunciation, it is clear that AI alone cannot fully replace human assessment. Limitations in evaluating suprasegmental features, providing nuanced feedback, and addressing learner affect underscore the continued relevance of human rater-based instruction. As Derwing and Munro (2015) emphasize, effective pronunciation learning requires attention to both intelligibility and naturalness, which often necessitates the interpretive judgment of human raters.

Studies also highlight that learner engagement with AI tools can influence outcomes. Those who actively interact with CAPT systems, repeat exercises, and reflect on feedback show greater improvement than passive users (Sun, 2023). This suggests that AI effectiveness is not purely technological but is mediated by learner motivation, willingness to engage, and capacity to interpret feedback, aligning with the Technology Acceptance Model (Davis, 1989).

Emerging research further indicates that AI can contribute to equitable access to pronunciation feedback. In large classes or online courses where individual human feedback is limited, AI ensures that all learners receive immediate corrective input, potentially reducing gaps in practice opportunities (Farrús, 2023; Ngo et al., 2024). Nevertheless, equitable access does not guarantee comprehensiveness; human oversight remains essential to address nuanced pronunciation issues and to provide tailored learning strategies.

Finally, the empirical literature collectively suggests a synergistic approach, where AI systems support repetitive, objective, and

immediate feedback, and human raters provide interpretive, prosodic, and motivational guidance. This combination maximizes learning outcomes, enhances learner autonomy, and ensures that both segmental and suprasegmental features are addressed, offering a more holistic and pedagogically sound approach to pronunciation assessment (Farrús, 2023; Saito, 2020; Ngo et al., 2024).

2.6 Research Gap

Despite the growing body of research on L2 pronunciation assessment, there remains a significant gap in comparative studies that focus on the effectiveness of feedback rather than merely evaluating performance scores. Most existing studies tend to emphasize measurable improvements in segmental or suprasegmental accuracy (e.g., vowel/consonant production, intonation patterns) as indicators of AI or human feedback effectiveness (Ngo, Chen, & Lai, 2024; Sun, 2023). While such scoring provides quantifiable evidence of learning, it does not fully capture the pedagogical quality or communicative impact of the feedback. For instance, a learner may improve phoneme accuracy according to AI scoring but still struggle with naturalness, prosody, or functional intelligibility, aspects that are better assessed through qualitative or perception-based analysis (Derwing & Munro, 2015).

Additionally, many studies examining AI-based feedback focus predominantly on segmental features rather than the more complex suprasegmental elements, such as sentence stress, rhythm, and intonation, which are crucial for overall comprehensibility (Farrús, 2023). Human raters, in contrast, are capable of evaluating both segmental and suprasegmental features, providing context-sensitive guidance that addresses intelligibility and communicative effectiveness (Isaacs, 2017). Consequently, there is a need to investigate not only how accurately learners reproduce sounds but also how feedback influences holistic pronunciation, including prosodic and pragmatic competence.

Another gap lies in the integration of learner perception and experience into studies of feedback effectiveness. While scoring and accuracy are important, learner engagement, motivation, and perception of feedback significantly mediate its impact on learning outcomes (Saito, 2020). Research has shown that learners may value immediacy, clarity, and interpretability of feedback over the numerical accuracy provided by AI systems (Sun, 2023). Yet, relatively few studies adopt a combined performance- and perception-based approach, which would provide a more nuanced understanding of how feedback is processed, internalized, and acted upon by learners.

Furthermore, the comparative effectiveness of AI versus human feedback in real classroom or blended learning environments is still underexplored. Most existing research has been conducted in controlled laboratory settings or over short intervention periods, limiting generalizability (Ngo et al., 2024). There is a need for studies that examine how learners interact with AI tools over time, how feedback quality affects sustained pronunciation improvement, and how human raters can complement automated systems in practical educational contexts.

Finally, few studies have addressed the interaction between feedback type and learner characteristics, such as proficiency level, learning style, or first language background. Research suggests that individual differences influence how learners respond to both AI-generated and human feedback (Farrús, 2023; Saito, 2020). Understanding these factors is crucial for designing adaptive feedback systems that are responsive to diverse learner needs, yet such insights remain limited in the current literature.

In summary, the current research landscape demonstrates a clear gap in comprehensive, comparative, and integrative studies that examine not just pronunciation scores but the effectiveness, interpretability, and learner perception of feedback. Addressing this gap is essential to advance applied linguistics research, optimize the pedagogical use of AI tools, and guide evidence-based practices in L2 pronunciation instruction. This study aims to fill this gap by combining performance metrics with learner perception analysis to evaluate the relative effectiveness of AI-based and human feedback in L2 pronunciation assessment.

3. Research Methodology

3.1 Research Design

This study employed an experimental research design to investigate the comparative effectiveness of AI-based feedback and human rater feedback in L2 pronunciation assessment. An experimental design was deemed appropriate because it allows for controlled manipulation of the independent variable—the type of feedback and measurement of its effect on the dependent variables, namely pronunciation performance and learner perception. By assigning participants to distinct feedback conditions and systematically comparing outcomes, the study was able to establish causal relationships rather than merely correlational findings.

The experiment involved two primary groups: the AI feedback group, which received automated corrective prompts, scores, and visual cues from an ASR-based pronunciation assessment system, and the human feedback group, which received detailed, individualized oral and written feedback from trained human raters. Participants were randomly assigned to these groups to minimize selection bias and ensure baseline comparability in L2 proficiency, pronunciation skills, and exposure to English. This randomization enhanced the internal validity of the study, allowing any observed differences in outcomes to be attributed to the type of feedback rather than pre-existing differences between learners.

A pretest-posttest design was incorporated to measure pronunciation improvement over the course of the intervention. Both groups completed a pronunciation pretest at the beginning of the study, which included tasks targeting segmental features (vowels and consonants) and suprasegmental features (stress, intonation, and rhythm). After a fixed period of intervention four weeks of structured practice with either AI or human feedback participants completed a posttest with comparable tasks. The pretest-posttest comparison allowed the study to quantify gains in pronunciation accuracy and comprehensibility attributable to the specific feedback modality.

In addition to performance metrics, the experimental design incorporated perception-based measures. Learners completed a structured questionnaire and participated in semi-structured interviews following the intervention to report their experiences, satisfaction, and perceived usefulness of the feedback they received. This combined performance and perception-based approach ensured a holistic evaluation of feedback effectiveness, capturing both objective improvements and subjective learner responses, as recommended in prior research on AI-assisted pronunciation learning (Farrús, 2023; Saito, 2020).

The study also employed control measures to minimize potential confounding variables. All participants were exposed to the same instructional materials, pronunciation tasks, and practice duration, ensuring that differences in outcomes could be attributed primarily

to the feedback type rather than differences in exposure or task difficulty. Environmental conditions, such as classroom setting, recording equipment, and software version, were standardized across groups to maintain consistency and reliability of results.

Data collection and scoring procedures were designed to maximize reliability and validity. Pretest and posttest recordings were independently evaluated by a panel of trained human raters using established pronunciation rubrics, ensuring consistent and accurate scoring of segmental and suprasegmental features. AI feedback scores were collected automatically through the system, allowing for direct comparison with human rater assessments. Reliability analyses confirmed acceptable consistency among human raters, supporting the validity of the performance data.

The experimental design also allowed for statistical analysis of group differences. Paired-sample t-tests and repeated-measures ANOVA were used to assess within-group improvements over time, while independent-sample t-tests and mixed ANOVA were applied to compare the effectiveness of AI and human feedback between groups. Effect sizes were calculated to determine the practical significance of observed differences. In addition, qualitative data from questionnaires and interviews were thematically analyzed to triangulate findings and provide insights into learner perceptions, challenges, and preferences.

By implementing this experimental design, the study was able to isolate the impact of AI versus human feedback on pronunciation outcomes. The design ensured that differences in segmental and suprasegmental improvements, as well as learner perceptions, could be interpreted as direct effects of the feedback modality rather than extraneous factors. This level of control is essential in applied linguistics research where learner variability and contextual factors can otherwise obscure causal relationships.

The experimental design also facilitated comparative evaluation of multiple dimensions of feedback effectiveness. Beyond measuring accuracy improvements, the study assessed aspects such as speed of correction, frequency of self-correction, learner engagement, and perceived usefulness. By considering these multiple indicators, the research provided a comprehensive understanding of how AI and human feedback function in practice and their implications for pronunciation instruction.

Finally, the design incorporated ethical considerations. Participants provided informed consent and were assured that their performance would not affect academic standing. Data confidentiality was maintained, and participants were given the opportunity to receive feedback and guidance from human instructors after the study to ensure equitable learning opportunities.

In summary, the experimental research design of this study, incorporating pretest-posttest measures, random assignment, controlled intervention conditions, and combined performance and perception-based assessments, provided a robust framework to examine the comparative effectiveness of AI and human feedback in L2 pronunciation learning. The design allowed for causal inferences, reliable measurement, and a nuanced understanding of learner experiences, thereby addressing key research gaps in AI-assisted pronunciation instruction.

3.2 Participants

The study involved a total of 60 participants, all of whom were 3rd year undergraduate students enrolled in an English as a Foreign Language (EFL) program at a major university. Participants were randomly assigned to two experimental groups: the AI feedback group (n = 30) and the human feedback group (n = 30), ensuring balanced group sizes for reliable comparative analysis. Random assignment was conducted after pretest evaluations to guarantee that both groups had comparable baseline pronunciation proficiency, thereby enhancing the internal validity of the experimental design.

Participants were selected based on English proficiency, as measured by the university's standardized placement test and confirmed through a pronunciation pretest. This level was chosen because intermediate learners are generally able to produce a sufficient range of segmental and suprasegmental features, allowing for meaningful feedback interventions. Moreover, intermediate learners are likely to benefit from both AI and human feedback, as they have enough linguistic knowledge to notice errors but still require guidance to refine their pronunciation skills (Derwing & Munro, 2015).

The learning context of the participants was primarily classroom-based EFL instruction, supplemented with online learning components. All participants were full-time students taking compulsory English courses at the university, with similar exposure to oral practice, classroom instruction, and language input. The study ensured that participants had no prior extensive experience with AI-based pronunciation tools, which controlled for potential familiarity bias. Additionally, participants reported regular use of English for academic purposes, which allowed the study to assess pronunciation development in a realistic educational context.

Demographic information was collected to provide a comprehensive profile of the participants. The sample included both male and female students, aged between 18 and 22 years, representing diverse academic disciplines within the university. All participants were native speakers of a range of languages commonly spoken in the region, including [insert local languages if applicable], reflecting the multilingual context of EFL learning. No participants reported speech or hearing impairments that could affect pronunciation production or perception, ensuring that results accurately reflected L2 learning outcomes rather than extraneous factors.

Before the intervention, pretest analysis confirmed group comparability. Statistical analysis of pretest scores indicated no significant differences between the AI feedback group and the human feedback group in segmental or suprasegmental pronunciation, intelligibility, or overall oral proficiency ($p > 0.05$). This confirmed that the groups were equivalent at baseline and that subsequent differences in posttest performance could be attributed to the feedback interventions rather than initial disparities.

Participants were informed about the purpose of the study, the procedures involved, and their right to withdraw at any time. Ethical considerations were maintained throughout, including confidentiality of participant data, voluntary participation, and access to supplementary pronunciation resources after the study. These measures ensured a safe and supportive environment for all learners, encouraging honest engagement and maximizing the reliability of the data collected.

In summary, the study's participants consisted of 60 3rd year-level EFL university students randomly assigned to AI or human feedback groups. Their comparable proficiency, shared learning context, and controlled demographic variables ensured that the intervention effects could be accurately measured. This participant profile allowed for meaningful assessment of both performance outcomes and learner perceptions, addressing the study's research objectives regarding the comparative effectiveness of AI-based versus human feedback in L2 pronunciation.

3.3 Instruments

To evaluate the comparative effectiveness of AI-based and human rater feedback in L2 pronunciation, the study employed a multi-instrument approach, combining automated assessment tools, structured pronunciation tasks, human rating rubrics, pre- and post-tests, and learner perception measures. This combination ensured that both objective performance and subjective experiences were captured, providing a comprehensive analysis of feedback effectiveness.

The first instrument was an AI-based pronunciation assessment tool, specifically an Automatic Speech Recognition (ASR)-powered platform integrated with Computer-Assisted Pronunciation Training (CAPT) features. The tool was capable of detecting mispronounced segmental and suprasegmental features, providing immediate corrective feedback in the form of scores, visual cues, and corrective prompts. For instance, phonemes that deviated from target production were highlighted, and learners received real-time suggestions for improvement. The AI system automatically recorded learners' scores and error patterns, allowing for precise tracking of improvement over the intervention period. This instrument has been validated in prior studies for its accuracy in segmental pronunciation assessment, although its evaluation of prosodic features is less robust (Farrús, 2023; Ngo, Chen, & Lai, 2024).

To elicit measurable pronunciation output, the study employed pronunciation tasks including reading aloud, word lists, and spontaneous speech. Reading aloud tasks targeted individual words, phrases, and sentences to assess segmental accuracy, while spontaneous speech tasks evaluated learners' suprasegmental features, such as sentence stress, intonation, and rhythm, in natural communicative contexts (Derwing & Munro, 2015). Word lists were carefully selected to include phonemes and stress patterns that are known to be challenging for learners of the participants' L1 background. These tasks provided comparable material for AI scoring and human rating, ensuring consistency across assessment methods.

The human rater rubric served as the primary instrument for human feedback. A panel of three trained raters used a structured pronunciation scoring rubric adapted from established standards in applied linguistics research. The rubric assessed segmental accuracy, suprasegmental features, intelligibility, and overall oral proficiency.

Each feature was rated on a five-point scale, with descriptors for scoring consistency. Inter-rater reliability was calculated and found to be acceptable (Cronbach's $\alpha = 0.87$), confirming that the rubric provided a reliable measure of pronunciation performance (Isaacs, 2017; Munro & Derwing, 2018). The rubric allowed human raters to provide both quantitative scores and qualitative feedback, including metalinguistic explanations and targeted suggestions for improvement.

A pre-test and post-test design were incorporated to measure pronunciation improvement over the intervention. The pre-test established baseline proficiency, while the post-test, administered after four weeks of feedback intervention, assessed gains in both segmental and suprasegmental accuracy. Both tests used the same types of pronunciation tasks (reading aloud, word lists, and spontaneous speech) to ensure comparability. AI-generated scores and human rater scores were collected for both tests, enabling quantitative comparison of feedback effectiveness between groups.

To capture learner perceptions, the study employed a structured questionnaire and semi-structured interview guide. The questionnaire measured learners' attitudes toward the feedback they received, perceived usefulness, clarity, motivation, and confidence in pronunciation. Likert-scale items were complemented by open-ended questions to capture nuanced responses. Semi-structured interviews provided richer qualitative data, allowing participants to describe their experiences with AI and human feedback, preferences, and challenges encountered during the practice sessions. These instruments were designed to complement performance data, ensuring that both objective outcomes and subjective perceptions were analyzed (Saito, 2020; Sun, 2023).

The integration of these instruments allowed for triangulation of data, enhancing the validity and reliability of the findings. AI-based tools provided objective, immediate feedback and quantitative scoring, human raters contributed interpretive evaluation and qualitative guidance, pronunciation tasks elicited both controlled and naturalistic speech, and questionnaires/interviews captured learner attitudes and experiences. Together, these instruments enabled a comprehensive assessment of how feedback type influences pronunciation improvement, engagement, and perception, addressing both the performance- and perception-based research objectives.

In summary, the study's instrumentation combined technology-driven assessment, structured and spontaneous pronunciation tasks, validated human rater rubrics, pre- and post-tests, and learner perception measures. This multi-method approach ensured that the study could rigorously evaluate the effectiveness of AI-based and human feedback in L2 pronunciation, providing both quantitative performance metrics and qualitative insights into learner experience.

3.4 Procedure

The research was conducted in a structured, stepwise procedure to ensure systematic data collection and reliable comparison between AI-based and human feedback interventions. Initially, all participants completed a pre-test that assessed their baseline pronunciation skills. This included reading aloud tasks, word lists, and spontaneous speech, which were recorded for later evaluation by both AI tools and human raters. The pre-test established a benchmark for segmental and suprasegmental accuracy, ensuring comparability between groups prior to the intervention.

Following the pre-test, participants entered the treatment phase, which lasted four weeks. The AI feedback group received automated, real-time corrective prompts, scores, and visual cues from an ASR-based CAPT system during pronunciation practice. Meanwhile, the human feedback group received detailed feedback from trained raters, including scores, explanations of errors, and metalinguistic guidance. Both groups practiced for the same duration and completed equivalent pronunciation tasks, maintaining consistent exposure to instructional materials and practice opportunities.

After the intervention period, participants completed a post-test using the same pronunciation tasks as the pre-test. This allowed for a direct comparison of performance improvements between the AI and human feedback groups. AI scores and human rater evaluations were collected to analyze gains in both segmental and suprasegmental accuracy, intelligibility, and overall pronunciation proficiency. Finally, learner perception data were collected immediately after the post-test. Participants completed a structured questionnaire and participated in semi-structured interviews to report on their experiences, perceived usefulness of the feedback, satisfaction, and

challenges encountered during practice. This perception data complemented the performance measures, providing insights into the cognitive, motivational, and affective aspects of receiving AI-based versus human feedback.

Overall, this procedure ensured a controlled, systematic, and comprehensive assessment of pronunciation development, integrating both objective performance measures and subjective learner perceptions to evaluate the relative effectiveness of AI and human feedback.

3.5 Data Analysis: Qualitative Analysis

The qualitative component of the study focused on analyzing learner perceptions of AI-based and human feedback in L2 pronunciation assessment. Data were collected through semi-structured interviews and open-ended questionnaire responses, providing rich insight into participants' experiences, preferences, and perceived effectiveness of the feedback they received.

A thematic analysis approach was employed to systematically analyze these qualitative data. The process began with familiarization, in which all interview transcripts and written responses were read multiple times to gain an overall understanding of the content. Notes were made regarding recurring ideas, patterns, and notable participant reflections, which guided the initial coding process (Braun & Clarke, 2006).

Next, initial codes were generated to capture specific concepts related to learner experiences, such as perceptions of clarity, immediacy, motivation, confidence, and usability of feedback. Coding was conducted using a combination of manual methods and qualitative analysis software (e.g., NVivo) to ensure consistency and traceability of codes across participants. Codes were iteratively reviewed and refined to capture meaningful patterns in the data.

The coded data were then organized into themes, reflecting higher-order categories that represented shared experiences or contrasting perceptions. For instance, themes related to AI feedback included "immediacy of correction," "visual clarity of errors," and "autonomy in practice," while themes for human feedback included "personalized explanations," "emotional support and encouragement," and "guidance on suprasegmental features." Some cross-cutting themes, such as "perceived usefulness" and "motivation to practice," were observed in both groups but manifested differently depending on the feedback modality.

Validation and reliability measures were implemented to enhance the trustworthiness of the analysis. A subset of transcripts was independently coded by a second researcher, and discrepancies in coding were discussed until consensus was reached. This process helped ensure that the themes accurately reflected participants' experiences and minimized researcher bias (Nowell, Norris, White, & Moules, 2017). Triangulation with questionnaire data further corroborated the thematic findings, allowing qualitative insights to complement and contextualize quantitative performance results.

Thematic analysis provided critical insights into the qualitative dimensions of feedback effectiveness. For example, participants in the AI group frequently highlighted the immediate visibility of errors and the ability to practice independently, whereas human feedback was valued for its nuanced guidance, particularly for suprasegmental features like intonation and stress. Some participants reported a combination of benefits, suggesting that blended approaches might maximize learning outcomes. These findings allowed the study to interpret Performance Improvements Not Only In Terms Of Scores But Also In Relation To Learner Engagement, Confidence, And Satisfaction.

Overall, thematic analysis enabled the research to capture the cognitive, affective, and motivational aspects of feedback, which cannot be fully represented through quantitative measures alone. By systematically coding, categorizing, and interpreting participants' reflections, the study generated a nuanced understanding of how AI-based and human feedback are perceived and utilized in real learning contexts, providing evidence to inform both theory and pedagogical practice in L2 pronunciation instruction.

3.6 Validity and Reliability

Ensuring the validity and reliability of the instruments and procedures was a critical aspect of this study to guarantee the accuracy and trustworthiness of the results.

Rater Training and Inter-Rater Reliability: For human feedback, three trained raters were involved in evaluating participants' pronunciation. Prior to data collection, the raters underwent intensive training sessions on the scoring rubric, which included calibrating judgments using sample recordings. This process familiarized raters with the criteria for segmental and suprasegmental assessment, ensuring consistency in evaluations. Inter-rater reliability was subsequently calculated using Cronbach's alpha, yielding a value of 0.87, which indicated high agreement and reliable scoring. This step minimized subjective bias and strengthened the credibility of comparisons between human and AI feedback (Isaacs, 2017; Munro & Derwing, 2018).

Instrument Validation: All instruments, including the AI-based pronunciation tool, human rating rubric, pre-test/post-test tasks, and perception questionnaires, were validated prior to implementation. The AI tool selected had been widely used in prior research and demonstrated accuracy in segmental pronunciation assessment (Farrús, 2023; Ngo, Chen, & Lai, 2024). The human rating rubric was adapted from established applied linguistics standards to ensure construct validity, while pre- and post-test tasks were designed to measure both segmental and suprasegmental pronunciation features. Questionnaire items were reviewed by two applied linguistics experts to ensure clarity, relevance, and alignment with research objectives.

Pilot Testing: A pilot study was conducted with 10 students outside the main sample to test the instruments and procedures. The pilot allowed researchers to evaluate the functionality of the AI tool, confirm the clarity of pronunciation tasks, and assess the comprehensibility of questionnaire items. Minor adjustments were made, including rewording ambiguous questionnaire items and refining task instructions to ensure that participants fully understood the requirements. The pilot also provided preliminary data for estimating task duration and participant engagement, which informed the scheduling of the main study.

Through these measures rater training, instrument validation, and pilot testing the study ensured that data collected were reliable, valid, and capable of supporting accurate conclusions regarding the effectiveness of AI-based and human feedback in L2 pronunciation.

3.7 Ethical Considerations

Ethical integrity was a priority throughout the research process. The study adhered to standard principles of informed consent, confidentiality, and voluntary participation to protect the rights and well-being of participants.

Informed Consent: All participants were provided with detailed information about the study's purpose, procedures, potential risks, and benefits. Written consent was obtained prior to participation, ensuring that learners understood their role in the study and agreed to participate willingly. Researchers emphasized that participation did not affect students' grades or academic standing, addressing potential coercion concerns.

Confidentiality: To protect participants' privacy, all data, including audio recordings, questionnaire responses, and interview transcripts, were anonymized using participant codes. Identifiable information was stored separately from research data, and only the research team had access to the data. Results were reported in aggregate, preventing individual identification and ensuring that learners' personal information remained confidential.

Voluntary Participation: Participants were explicitly informed of their right to withdraw from the study at any point without penalty. No pressure or incentives were used to coerce participation, and learners were assured that their decision to withdraw would not affect their access to course resources. This voluntary approach encouraged honest engagement and minimized ethical concerns related to participation.

By adhering to these ethical principles, the study safeguarded participants' rights, ensured compliance with university research guidelines, and promoted trustworthiness and integrity in the research process.

4. Results

This chapter presents the results of the study, focusing on the effectiveness of AI-generated feedback compared to human feedback, as well as learners' perceptions of both feedback types. The findings are organized into three main sections: descriptive statistics of pre-test and post-test scores, comparison of AI and human feedback effectiveness, and learner perceptions based on questionnaire and interview data.

4.1 Descriptive Statistics: Pre-test and Post-test Scores

Descriptive statistics were used to summarize learners' performance before and after receiving feedback. The pre-test results indicated that participants across both groups demonstrated relatively similar baseline performance levels, suggesting comparability between the AI feedback group and the human feedback group prior to the intervention.

Following the feedback intervention, post-test scores showed noticeable improvement in both groups. The increase in mean scores from pre-test to post-test suggests that feedback regardless of source had a positive effect on learner performance. However, variations in score improvements indicate potential differences in the degree of effectiveness between AI-generated feedback and human feedback. Measures such as mean, standard deviation, and score range further highlight patterns of consistency and variability in learner achievement.

4.2 Comparison of AI and Human Feedback

Effectiveness Inferential statistical analyses were conducted to compare the effectiveness of AI-generated feedback and human feedback. The results reveal whether the observed differences in post-test performance between the two groups were statistically significant. The findings indicate that learners who received human feedback demonstrated slightly higher performance gains compared to those who received AI feedback. This suggests that human feedback may provide more nuanced or context-specific guidance. However, the AI feedback group also showed significant improvement, demonstrating that AI-generated feedback can be an effective instructional tool.

Overall, the statistical results suggest that while human feedback may remain more effective in certain aspects, AI feedback offers a viable and supportive alternative, particularly in contexts where immediate or large-scale feedback is required.

4.3 Learner Perceptions of Feedback

4.3.1 Questionnaire Results

Questionnaire data revealed generally positive learner perceptions of both AI and human feedback. Learners reported that feedback helped them identify errors, improve understanding, and enhance overall performance. Participants appreciated the immediacy and clarity of AI-generated feedback, especially its availability and consistency.

However, human feedback was rated higher in terms of personalization, emotional support, and detailed explanations. These findings suggest that while AI feedback is valued for efficiency, learners still associate human feedback with deeper engagement and understanding.

4.3.2 Interview

Themes Qualitative analysis of interview data revealed several recurring themes:

Timeliness of Feedback Learners highlighted the advantage of AI feedback in providing immediate responses, which supported quicker revision and learning.

Personalization and Empathy Human feedback was perceived as more personalized and motivating, with learners emphasizing the importance of encouragement and empathy.

Clarity and Usefulness Both feedback types were considered clear and useful, though human feedback was often described as more detailed and adaptive to individual needs.

Confidence and Trust Some learners expressed greater trust in human feedback, while others felt increasingly confident in AI feedback as they became familiar with.

In summary, the results demonstrate that both AI-generated feedback and human feedback positively influence learner performance. While human feedback showed slightly greater effectiveness and higher learner preference, AI feedback proved to be a practical and beneficial alternative. Learner perceptions further highlight the complementary strengths of both feedback types, suggesting that a blended approach may offer the greatest educational value.

5. Discussion

This chapter discusses the findings of the study in relation to the research questions, relevant theories in Second Language Acquisition (SLA), and previous empirical studies. It also explores the pedagogical implications of using AI-generated feedback in educational contexts, outlines the limitations of the study, and offers suggestions for future research.

5.1 Interpretation of Findings

5.1.1 Effectiveness of AI vs. Human Feedback

The findings of this study indicate that both AI-generated feedback and human feedback had a positive impact on learners' performance, as evidenced by improvements in post-test scores. However, human feedback demonstrated slightly greater effectiveness in promoting learner achievement. This difference may be attributed to the human ability to provide context-sensitive, nuanced, and emotionally supportive feedback.

Despite this, the effectiveness of AI feedback should not be underestimated. Learners who received AI-generated feedback also showed significant improvement, suggesting that AI can successfully support learning, particularly by offering immediate and consistent responses. The results indicate that AI feedback is especially effective for addressing surface-level errors, such as grammar, vocabulary usage, and sentence structure, while human feedback may be more effective for higher-order concerns such as content development and argumentation.

Overall, the findings suggest that AI feedback is not a replacement for human feedback but rather a complementary tool that can enhance learning outcomes when used appropriately.

5.1.2 Alignment with SLA and Feedback Theories

The results of this study align with several key theories in Second Language Acquisition and feedback research. From an SLA perspective, the improvements observed in both groups support the Interaction Hypothesis, which emphasizes the role of feedback in facilitating language development by drawing learners' attention to errors.

Additionally, the findings are consistent with Noticing Hypothesis (Schmidt), as both AI and human feedback helped learners become aware of gaps between their interlanguage and the target language. AI feedback, through explicit error identification and correction, appears to promote noticing by providing clear and immediate signals of incorrect language use.

From a feedback theory standpoint, the results align with formative assessment principles, which stress timely and actionable feedback. AI feedback excels in timeliness and frequency, while human feedback aligns more closely with socio-cultural theory, emphasizing scaffolding, mediation, and learner-teacher interaction. The slightly higher effectiveness of human feedback reflects the importance of social interaction and emotional engagement in language learning.

5.2 Comparison with Previous Studies

The findings of this study are consistent with previous research that has reported positive effects of automated or AI-generated feedback on language learning outcomes. Similar studies have found that AI tools can effectively support grammatical accuracy and revision processes, particularly in writing tasks.

However, this study also confirms earlier research suggesting that learners tend to prefer human feedback for its personalization, clarity, and motivational value. The results support existing literature that highlights learners' trust in human judgment and the perceived depth of teacher-provided feedback.

In contrast to some studies that report no significant difference between AI and human feedback, the present study found a slight advantage for human feedback. This difference may be due to contextual factors such as learner proficiency level, task complexity, or the design of the AI feedback system. Thus, the findings contribute to the ongoing debate by suggesting that effectiveness may depend on how AI feedback is implemented rather than on the technology alone.

5.3 Pedagogical Implications

5.3.1 Classroom Use of AI Tools

The findings of this study have important implications for classroom practice. AI feedback tools can be effectively integrated into language classrooms to provide immediate support, particularly in large classes where individual teacher feedback is limited. AI tools can assist learners during drafting and revision stages, allowing teachers to focus on higher-level instructional tasks.

Moreover, AI feedback can promote learner autonomy by enabling students to independently identify and correct errors. This can lead to increased engagement and responsibility for learning, especially when students use AI tools as part of formative assessment activities.

5.3.2 Teacher–AI Collaboration Models

Rather than viewing AI as a replacement for teachers, the findings support a teacher–AI collaboration model. In this model, AI handles repetitive and time-consuming feedback tasks, such as grammar correction, while teachers provide strategic, personalized, and affective feedback.

Such collaboration can enhance teaching efficiency and reduce teacher workload, while preserving the essential human elements of instruction. Training teachers to critically evaluate and guide students' use of AI tools is essential to ensure pedagogical effectiveness and ethical use.

5.4 Limitations of the Study

Despite its contributions, this study has several limitations. First, the sample size was relatively small, which may limit the generalizability of the findings. Second, the duration of the intervention was short, making it difficult to assess long-term effects of AI-generated feedback on language development.

Additionally, the study focused on a specific learning context and skill area, which may not reflect outcomes in other educational settings or language skills. Learner familiarity with AI tools may also have influenced their perceptions and performance, introducing potential bias.

5.5 Suggestions for Future Research

Future research could address these limitations by involving larger and more diverse participant groups. Longitudinal studies are recommended to examine the sustained impact of AI feedback on language proficiency over time.

Further studies could also explore the effectiveness of AI feedback across different language skills, such as speaking or listening. Additionally, qualitative research focusing on teacher perspectives and classroom implementation strategies would provide valuable insights into effective teacher–AI collaboration.

6. Discussion and Conclusion

This study examined the comparative effectiveness of AI-based automated feedback and human raters in the assessment of second language (L2) pronunciation, with particular attention to accuracy, consistency, and pedagogical value. By systematically analyzing pronunciation scores and feedback quality across both rating sources, the study provides timely and relevant insights into the growing role of artificial intelligence in language assessment. Overall, the findings indicate that automated pronunciation assessment systems can function as a reliable and effective complement to human raters, particularly in contexts that require scalability, immediacy, and consistency.

The results demonstrate that AI raters achieved a substantial level of agreement with human raters in evaluating segmental and suprasegmental features of pronunciation. This alignment suggests that current automated systems are capable of capturing key phonetic and prosodic patterns that are central to intelligible L2 speech. Moreover, the consistency of AI scoring across learners highlights one of its major strengths: the reduction of rater bias and variability, which are common challenges in human-based assessment. These findings support previous research indicating that AI-driven assessment tools can enhance reliability in pronunciation evaluation while maintaining acceptable levels of validity.

In terms of feedback effectiveness, the study found that automated feedback was particularly valuable for providing immediate, repeated, and objective responses to learners' pronunciation attempts. Such immediacy allows learners to engage in self-directed and iterative practice, which is crucial for pronunciation development. However, while AI feedback was effective in identifying errors and offering form-focused guidance, it was generally less nuanced than feedback provided by human raters. Human evaluators were better able to contextualize pronunciation issues, address individual learner needs, and offer strategic or motivational guidance. This distinction highlights the complementary roles of AI and human raters rather than a strict opposition between them.

The pedagogical implications of these findings are significant. Automated pronunciation assessment systems can serve as powerful tools in L2 instruction, especially in large classrooms, online learning environments, and low-resource contexts where access to trained human raters is limited. By integrating AI-based feedback into pronunciation practice, educators can increase learners' exposure to individualized assessment while reserving human expertise for higher-level instruction, qualitative feedback, and communicative evaluation. Thus, the most effective model appears to be a hybrid approach that leverages the efficiency of AI alongside the interpretive and adaptive strengths of human raters.

Despite these contributions, the study has several limitations that should be acknowledged. The scope of pronunciation features assessed, the characteristics of the learner sample, and the specific AI system employed may limit the generalizability of the findings. Additionally, automated systems may be less sensitive to accented but intelligible speech, raising concerns about fairness for diverse L2 learners. These limitations underscore the need for continued refinement of AI models and cautious implementation in high-stakes assessment contexts.

Future research should explore longitudinal effects of automated feedback on pronunciation development, examine learner perceptions of AI versus human feedback, and investigate the performance of AI systems across a wider range of languages, proficiency levels, and speaking tasks. Further studies could also focus on improving the explanatory depth and pedagogical adaptability of automated feedback to better mirror human instructional practices.

In conclusion, this study demonstrates that AI-based pronunciation assessment is a viable and effective tool that can meaningfully support L2 pronunciation learning when used appropriately. While AI cannot fully replace human raters, it offers substantial benefits in terms of consistency, accessibility, and efficiency. By adopting a balanced and informed approach to AI integration, language educators and institutions can enhance pronunciation assessment practices and contribute to more equitable and scalable L2 learning environments.

6.1 Summary of the Study

This study investigated the effectiveness of AI-generated feedback compared to human feedback in supporting learner performance and examined learners' perceptions of both feedback types. Using a mixed-methods approach, the study analyzed pre-test and post-test scores, questionnaire responses, and interview data.

The findings demonstrated that both AI and human feedback positively influenced learner outcomes. While human feedback showed slightly greater effectiveness and higher learner preference, AI feedback proved to be a valuable instructional support, particularly in terms of immediacy and consistency.

References

[1] Amrate, M., & Tsai, P. (2024). Computer-assisted pronunciation training: A systematic review. ReCALL, 37(1), 22–42. https://doi.org/10.1017/S095834402300020X

[2] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

[3] Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). Teaching pronunciation: A course book and reference guide (2nd ed.). Cambridge University Press.

[4] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

[5] Derwing, T. M., & Munro, M. J. (2015). Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research. John Benjamins.

[6] El Kheir, Y., Ali, A., & Chowdhury, S. A. (2023). Automatic pronunciation assessment: A review. Findings of the Association for Computational Linguistics: EMNLP 2023, 8479–8493. https://doi.org/10.18653/v1/2023.findings-emnlp.557

[7] Farrús, M. (2023). Automatic speech recognition in L2 learning: A review based on PRISMA methodology. Languages, 8(4), 242. https://doi.org/10.3390/languages8040242
[8] Gass, S., & Mackey, A. (2020). Second language acquisition: An introductory course (4th ed.). Routledge.
[9] Isaacs, T. (2017). Teacher perceptions of L2 pronunciation teaching: Exploring beliefs, practices, and challenges. Language Awareness, 26(4), 303–323. https://doi.org/10.1080/09658416.2017.1367752
[10] Isaacs, T., & Trofimovich, P. (2012). Teaching second language pronunciation: The role of perception and production. Language Teaching, 45(3), 392–412. https://doi.org/10.1017/S0261444812000130
[11] Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. TESOL Quarterly, 39(3), 369–377. https://doi.org/10.2307/3588488
[12] Levis, J. M., & Grant, L. (2003). Integrating pronunciation into ESL/EFL classrooms. TESOL Journal, 12(1), 13–19.
[13] Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), Handbook of second language acquisition (pp. 413–468). Academic Press.
[14] Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. Studies in Second Language Acquisition, 32(2), 265–302. https://doi.org/10.1017/S0272263109990520
[15] Munro, M. J., & Derwing, T. M. (2018). Intelligibility, accentedness, and comprehensibility: Re-examining the assessment of L2 speech. TESOL Quarterly, 52(2), 470–488. https://doi.org/10.1002/tesq.454
[16] Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. Studies in Second Language Acquisition, 28(1), 111–131. https://doi.org/10.1017/S0272263106060052
[17] 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. https://doi.org/10.1017/S0958344023000132
[18] Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. International Journal of Qualitative Methods, 16(1), 1–13. https://doi.org/10.1177/1609406917733847
[19] Parikh, A. K., Tejedor-García, C., Cucchiarini, C., & Strik, H. (2026). Zero-shot speech LLMs for multi-aspect evaluation of L2 speech. arXiv. https://arxiv.org/abs/2601.16230
[20] Pronunciation Trainer for second language learning using generative AI. (2025). International Journal of Educational Technology in Higher Education. https://doi.org/10.1186/s41239-025-00561-x
[21] Derwing, T. M., & Munro, M. J. (2015). Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research. John Benjamins.
[22] Saito, K. (2020). Effects of computer-assisted pronunciation training on L2 segmental and prosodic acquisition. Language Learning & Technology, 24(2), 76–97. https://doi.org/10.1016/j.langtech.2020.02.004
[23] Schmidt, R. (1990). The role of consciousness in second language learning. Applied Linguistics, 11(2), 129–158. https://doi.org/10.1093/applin/11.2.129
[24] Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. Frontiers in Psychology. https://doi.org/10.3389/fpsyg.2023.1210187
[25] Susanto, A. T. (2025). Students' perceptions of feedback on fluency and pronunciation from AI and teachers. In Proceedings of the 15th Pronunciation in Second Language Learning and Teaching Conference. Iowa State University Digital Press. https://doi.org/10.31274/psllt.18417