Review of Applied Science and Technology
Volume 04, Issue 02 (2025)
Page No: 473 – 500
Doi: 10.63125/1wyj3p84

THE IMPACT OF HUMAN-MACHINE INTERACTION ON ENGLISH PRONUNCIATION AND FLUENCY: CASE STUDIES USING AI SPEECH ASSISTANTS

Elmoon Akhter[1]

1 *MA in English, State University of Bangladesh, Dhaka, Bangladesh.*
Email: elmoonshimu@gmail.com

Abstract

This systematic review investigates the impact of human-machine interaction, specifically through AI-powered speech assistants, on the development of English pronunciation and fluency among second language learners. Drawing on 54 peer-reviewed studies published between 2005 and 2023, the review synthesizes findings from experimental, quasi-experimental, longitudinal, and mixed-method research that explore how tools such as Siri, Alexa, Google Assistant, and ELSA Speak influence oral language performance. The review follows the PRISMA 2020 guidelines to ensure methodological transparency, with eligibility criteria focusing on studies that employed measurable outcomes related to segmental and suprasegmental pronunciation features, temporal fluency metrics, acoustic analysis, and learner perceptions. The results reveal consistent improvements in phoneme articulation, word stress, intonation, speech rate, and pause reduction after sustained AI-mediated practice. Learners also demonstrated increased metacognitive awareness, greater confidence, and behavioral shifts toward self-regulated learning, particularly when interpreting implicit feedback from AI systems. Furthermore, the adaptability of AI speech tools across cultural contexts and learning settings highlights their scalability and pedagogical value. This review affirms the growing role of AI speech assistants as effective, accessible, and motivational tools for enhancing spoken English proficiency and offers recommendations for their integration into second language instruction and learner autonomy frameworks.

Keywords
AI Speech Assistants; English Pronunciation; Second Language Acquisition; Human-Machine Interaction; Fluency Development

Citation:
Akhter, E. (2025). *The impact of human-machine interaction on English pronunciation and fluency: Case studies using AI speech assistants. Review of Applied Science and Technology*, 4(2), 473–500. https://doi.org/10.63125/1wyj3p84

Received:
April 20, 2025

Revised:
May 14, 2025

Accepted:
June18, 2025

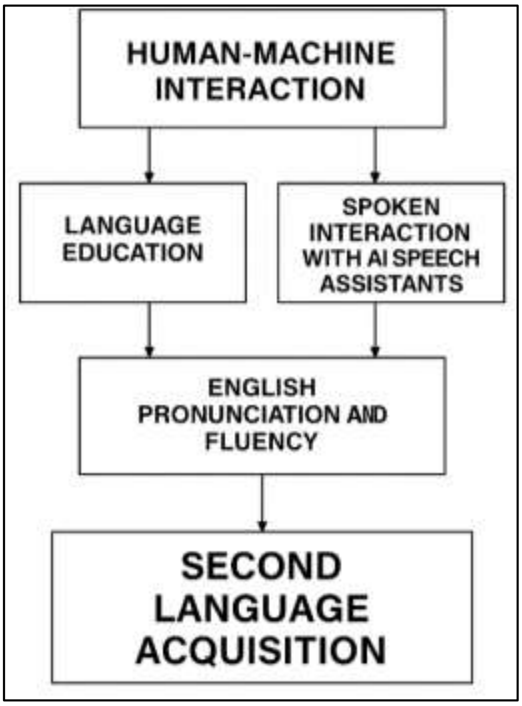
Published:
July 25, 2025

Review of Applied Science and Technology
Volume 04, Issue 02 (2025)
Page No: 473 – 500
Doi: 10.63125/1wyj3p84

INTRODUCTION

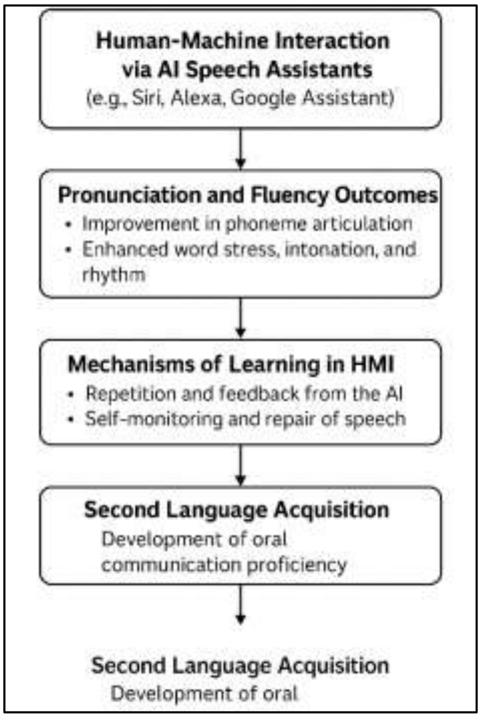
Human-Machine Interaction (HMI), often referred to as Human-Computer Interaction (HCI) when focused on computer-based systems, describes the study and design of interaction between people (users) and machines (computers, AI agents, and other digital systems) (Zhu et al., 2020). Within the context of language education, HMI serves as the foundation for AI-mediated learning environments where intelligent agents interact with users in dialogic and responsive ways (Zhao et al., 2020). A subset of HMI, spoken interaction with AI speech assistants, forms the core mechanism of this study. Speech assistants, such as Amazon's Alexa, Apple's Siri, and Google Assistant, utilize natural language processing (NLP), machine learning, and speech recognition technologies to respond to voice commands and engage in rudimentary conversations (Xiong et al., 2020). English pronunciation refers to the articulation and phonological accuracy of sounds, intonation, and stress patterns in spoken English, while fluency involves the fluidity, speed, and coherence of speech production without undue pauses or hesitations (Wen et al., 2020). These linguistic competencies are central to second language acquisition (SLA), particularly in oral communication proficiency. Globally, English serves as a lingua franca, functioning as a primary or secondary language in diplomatic, academic, economic, and technological domains (Tan et al., 2020). This international status has driven widespread initiatives to enhance English education, especially oral proficiency, across non-native contexts (Yao et al., 2019). Pronunciation and fluency remain among the most challenging areas for learners, primarily due to the limited exposure to authentic English speech models, varied phonotactic structures across first languages, and insufficient opportunities for spontaneous practice (Wang et al., 2019). Traditional classroom approaches to pronunciation instruction often rely on repetitive drills and teacher modeling, which may not fully address the real-time, interactive, and context-sensitive nature of spoken communication (Shi & Lee, 2019). As a result, alternative tools and technologies are being explored to provide learners with more immersive, adaptive, and autonomous ways to practice spoken English (Wang et al., 2018). The rise of AI speech assistants represents one such technological development, offering scalable, always-available interlocutors capable of real-time interaction.

Figure 1: Role of Human-Machine Interaction via AI Speech Assistant



The use of AI speech assistants in language learning has evolved from auxiliary roles in vocabulary support and grammar explanations to active participation in oral language practice (Cao et al., 2018). By enabling conversational exchanges through voice recognition and synthesis, these technologies mimic human interlocutors in meaningful ways. Research on mobile-assisted language learning (MALL) has shown that such tools increase learners' willingness to communicate, reduce anxiety, and offer immediate feedback in pronunciation tasks (He et al., 2017). For instance, studies on Google Assistant usage among EFL learners revealed improvements in segmental features of pronunciation, such as vowel clarity and consonant production, along with enhancements in prosodic elements like intonation and rhythm (Yao et al., 2019). Similarly, voice-interactive agents have been reported to boost learners' speaking confidence and autonomy in out-of-class contexts, thereby enriching the overall language learning experience. These outcomes are especially pertinent in countries with limited access to native speakers or qualified pronunciation instructors, where speech assistants can offer a cost-effective, widely deployable alternative (Wang et al., 2019).

Figure 2: Human-Machine Interaction via AI Speech Assistants



Empirical investigations into HMI's linguistic effects have emphasized the cognitive and perceptual dimensions of pronunciation development. Neural studies indicate that repeated interaction with consistent phonetic input can aid in establishing robust auditory targets for L2 learners (He et al., 2017). AI assistants offer consistent and repeatable pronunciation models, which contrasts with the variable and often imperfect output from peer interlocutors in traditional classrooms (McCrocklin, 2016). In addition, feedback from AI assistants—whether implicit through repetition or explicit through correction—has been found to support learners' phonological awareness. The role of input enhancement and negotiation of meaning, long considered central to SLA, is operationalized through these devices as learners adjust their speech to be recognized correctly by the assistant, thereby modifying pronunciation to increase intelligibility (Cao et al., 2018). Studies involving Siri and Alexa demonstrate that learners engage in such self-monitoring and repair behaviors, leading to more refined and fluent speech over time (Wang et al., 2018). In addition, Cross-cultural case studies on AI-mediated pronunciation training offer further insights into the socio-educational variables affecting HMI outcomes. For instance, research in East Asian contexts highlights the role of AI speech assistants in addressing L1-related pronunciation issues, such as vowel reduction and syllable timing in Korean and Mandarin speakers. In Middle Eastern regions, where cultural norms may restrict frequent spoken English interaction, speech assistants have been used as private, stigma-free tools for oral practice (Cao et al., 2018). Latin American and European studies reveal similar patterns, where learners report using assistants to simulate dialogues, repeat problematic phrases, and receive indirect correction through speech recognition errors (He et al., 2017). These findings underscore the global applicability of AI speech interfaces in addressing diverse phonological challenges across learner populations. Furthermore, longitudinal data suggests that frequent interaction with these tools correlates with measurable gains in fluency metrics, including speech rate, mean length of utterance, and reduction in filler use.

In terms of design and pedagogical affordances, AI speech assistants operate as both tools and environments for language learning. They facilitate interactionist SLA principles by creating authentic turn-taking exchanges and promoting negotiation of meaning (MacKay & Flege, 2004). Their multimodal capabilities—such as integrating speech with textual prompts, visual confirmations, or follow-up questions—support varied learning styles and enhance engagement (Hughes, 2004). Studies also report that learners develop metacognitive strategies while interacting with speech assistants, such as self-scaffolding, repetition monitoring, and paraphrasing to increase recognition accuracy. These interactions occur across informal and formal contexts, including smart classrooms, mobile learning settings, and home environments. The portability and familiarity of these technologies, especially among younger learners, have contributed to their increasing integration into language education programs. Additionally, research indicates that such tools can bridge the gap between in-class instruction and out-of-class practice, which is critical for sustained pronunciation development (Murphy & Baker, 2015).

Research methodologies employed in studying HMI and AI speech assistants vary widely, encompassing experimental, ethnographic, and mixed-methods designs. Acoustic analysis using software like PRAAT, speech recognition logs, self-report questionnaires, and expert ratings are commonly employed to evaluate gains in pronunciation and fluency. Case studies—particularly longitudinal ones—provide rich accounts of learners' adaptive strategies, error correction behaviors, and evolving interaction patterns with AI assistants. These studies often triangulate data from learner speech samples, system feedback, and participant reflections to map the progression of spoken language skills. Additionally, comparative analyses between traditional methods (e.g., instructor-led drills) and AI-mediated tasks reveal that the latter often result in higher learner engagement, retention of corrected forms, and transfer of learning to spontaneous speech(Chapelle, 2009). Such empirical work forms the basis for understanding the nuanced impact of human-machine interaction on second language pronunciation and fluency acquisition.The primary objective of this study is to investigate the impact of interactive engagement with AI-powered speech assistants on the pronunciation accuracy and oral fluency of English language learners across diverse linguistic and educational backgrounds. This investigation focuses on capturing the transformation of learners' spoken language abilities through structured, longitudinal interaction with speech-based AI tools such as virtual assistants that employ voice recognition and synthesized responses.

Specifically, the study aims to assess how repeated verbal interactions with these technologies contribute to segmental and suprasegmental improvements in pronunciation, including clarity of vowels and consonants, stress patterns, intonation, and rhythm. The research also seeks to evaluate gains in fluency, measured by speech rate, mean length of utterance, reduction in pauses, and self-repair patterns. By conducting multiple case studies, the study endeavors to provide granular insights into individual learner trajectories, identifying the patterns, challenges, and benefits associated with their engagement with AI speech systems over a sustained period. Furthermore, the research aims to document how learners adapt their speech to improve intelligibility when interacting with AI systems that offer indirect feedback mechanisms, such as speech recognition errors or reformulations. The objective includes exploring how these technologies function as both practice tools and communicative partners, fostering autonomy and learner-initiated correction. The study also examines learner perceptions, comfort levels, and self-assessed progress, adding a qualitative dimension to the analysis. By encompassing both quantitative measures of linguistic improvement and qualitative accounts of user experience, the study aims to deliver a comprehensive account of how human-machine spoken interaction supports language development in real-world and educational contexts.

LITERATURE REVIEW

The integration of AI speech assistants into language learning environments has emerged as a pivotal development in educational technology, particularly in enhancing spoken English proficiency among non-native speakers. This literature review examines the theoretical and empirical foundations underlying the use of human-machine interaction to support pronunciation and fluency development. It outlines key contributions from fields such as applied linguistics, second language acquisition, educational technology, human-computer interaction, and speech processing. Central to the discourse are the mechanisms by which AI systems influence learners' articulation patterns, prosodic control, and verbal fluidity. The review categorizes the literature thematically, beginning with foundational theories of pronunciation and fluency in second language acquisition, moving into the capabilities and pedagogical affordances of AI speech assistants, and culminating in detailed accounts of experimental studies and cross-cultural implementations. This structured synthesis provides a comprehensive understanding of current knowledge, identifies research gaps, and contextualizes the significance of interactive AI tools in developing real-time spoken language skills.

Pronunciation and Fluency in Second Language Acquisition

Pronunciation in second language acquisition encompasses both segmental and suprasegmental features, which together shape intelligibility, comprehensibility, and accentedness in learner speech. Segmental aspects refer to individual phonemes—consonants and vowels—while suprasegmental features include rhythm, stress, intonation, and pitch. These elements jointly contribute to the perception of fluency and clarity in spoken interaction. Research has consistently emphasized the importance of intelligible pronunciation over native-like accuracy, particularly in multilingual communicative contexts (Derwing & Munro, 2005). Learners often struggle with phonemic contrasts that are absent in their first language, resulting in misperceptions or misarticulations that hinder effective communication. Phonological transfer from the L1 exerts a strong influence on how learners encode and produce new sounds in the L2, creating persistent challenges for learners from diverse linguistic backgrounds. Studies show that adult learners retain a strong influence of their native phonological system due to entrenched auditory templates, which makes accurate production of novel phonemes difficult (Dlaska & Krekeler, 2008). Instructional approaches have focused on explicit phonetic training, auditory discrimination tasks, and articulatory feedback to target pronunciation development. Tools such as visual spectrograms and slow speech modeling have been employed to support learners' awareness and control over specific phonetic features. The integration of pronunciation instruction into communicative language teaching has received renewed interest as scholars seek to balance form-focused accuracy with meaningful interaction. Recent pedagogical frameworks emphasize the role of repetition, modeling, and intelligibility-driven feedback in shaping learners' phonological development. These approaches align with findings from acoustic studies that demonstrate how targeted exposure and deliberate practice can recalibrate learners' perceptual categories over time (Sardegna et al., 2017).

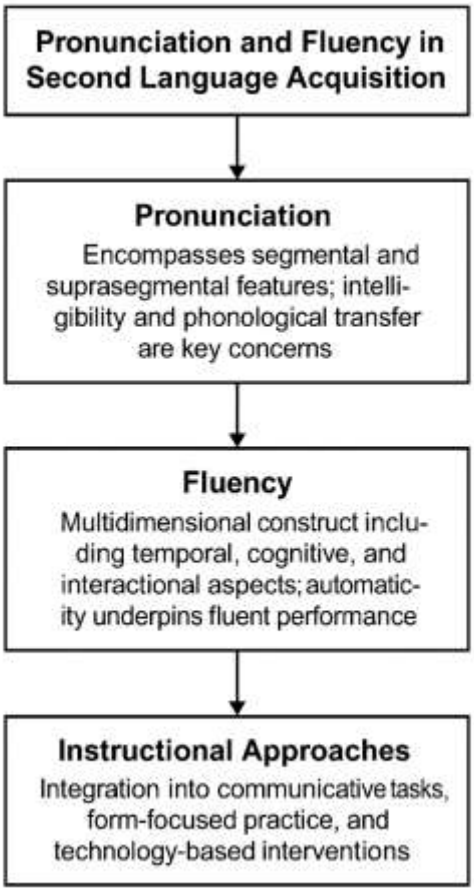
Oral fluency, though often associated with speaking speed, represents a multidimensional construct that includes temporal, cognitive, and interactional aspects of speech production. Temporal fluency involves speech rate, pause frequency, and mean length of utterance, while cognitive fluency reflects processing ease and lexical retrieval efficiency during spontaneous speech. Interactional fluency focuses on turn-taking behavior and responsiveness in dialogues. Researchers have proposed a range of metrics for fluency assessment, including syllables per minute, articulation rate, and frequency of filled and unfilled pauses (Sardegna et al., 2017). These measures offer insights into the psycholinguistic and discourse-level functioning of learners as they navigate second language speech. Studies reveal that fluency is highly sensitive to task type, familiarity, and learner proficiency, with narrative retelling, picture description, and role-plays often used as elicitation tools (Dlaska & Krekeler, 2013). While speech rate is frequently used as a primary indicator, it does not necessarily correlate with communicative competence unless accompanied by syntactic and lexical complexity. Cognitive models of fluency suggest that automaticity, defined as the ability to retrieve and articulate language forms with minimal conscious effort, underpins fluent performance (Moyer, 2014). Automaticity is influenced by repeated practice, exposure to formulaic sequences, and reduced attentional load during speech processing. Studies have also examined the role of working memory and attention control in supporting fluent speech, especially in time-pressured conditions. Researchers have differentiated between breakdown fluency, which measures the extent of hesitations and false starts, and repair fluency, which involves self-correction strategies and restructuring. Findings indicate that experienced learners deploy more efficient repair mechanisms and exhibit fewer pauses compared to novices, highlighting the role of strategic competence in oral fluency development.

Pronunciation and fluency acquisition are shaped not only by linguistic input but also by cognitive capacities and affective variables. Cognitive dimensions such as phonological working memory, auditory discrimination skills, and proceduralization of articulatory routines significantly influence

Review of Applied Science and Technology
Volume 04, Issue 02 (2025)
Page No: 473 – 500
Doi: 10.63125/1wyj3p84

learners' ability to acquire and produce accurate speech sounds (LeVelle & Levis, 2014). Learners with stronger phonological memory tend to show greater gains in pronunciation tasks and retain new phonetic patterns more effectively over time. Neurocognitive studies have demonstrated that L2 pronunciation learning involves the activation of auditory, motor, and speech planning regions of the brain, with neural plasticity facilitating adaptation to unfamiliar sound patterns through repeated exposure (Jenkins, 2002). The automatization of phonological production contributes to increased fluency, allowing learners to allocate cognitive resources to higher-level discourse planning rather than articulation. In addition to cognitive mechanisms, affective factors such as anxiety, motivation, and self-perceived speaking ability play pivotal roles. Language anxiety has been shown to negatively affect pronunciation accuracy and fluency by disrupting speech planning and increasing the frequency of hesitations and self-corrections. Learners who perceive themselves as poor speakers often avoid speaking opportunities, limiting the practice necessary for improvement. Conversely, highly motivated learners are more likely to engage in extensive speaking practice, monitor their speech output, and seek corrective feedback, all of which contribute to pronunciation gains. Studies also indicate that learners' attitudes toward accented speech affect their willingness to adopt target-like pronunciation norms, with some embracing a global English perspective that values intelligibility over native-like accuracy (Kibishi et al., 2014). Emotional resilience and positive learning environments contribute to greater risk-taking in speaking tasks, which facilitates experimentation with new phonetic forms and increases fluency development.

Figure 3: Pronunciation and Fluency Development in Second Language Acquisition



Instructional approaches to pronunciation and fluency in second language classrooms have evolved from mechanical drilling and imitation to more communicative and technology-enhanced frameworks. Traditional methods focused on phonetic transcription, minimal pairs, and repetition drills, which often lacked meaningful communicative contexts. Contemporary pedagogy emphasizes the integration of pronunciation instruction into communicative tasks that promote both accuracy and fluency, allowing learners to practice speech features in authentic scenarios. Form-focused instruction, combined with interactional practice, has been shown to support long-term gains in both segmental and prosodic accuracy (Lasi, 2020). Explicit instruction that draws learners' attention to problematic phonemes and offers articulatory feedback has been linked to significant improvements in intelligibility and comprehensibility. Technology-based interventions, including computer-assisted pronunciation training (CAPT), mobile applications, and AI-integrated systems, have become prominent in recent years (Foote et al., 2012). These tools provide learners with individualized practice, immediate feedback, and multisensory input, thereby supporting the development of pronunciation and fluency outside the classroom. Visual feedback through waveform displays and spectrograms helps learners understand acoustic properties of speech and refine their articulation. Speech recognition software has enabled learners to receive automated evaluation of their spoken output, promoting self-monitoring and adjustment. Studies show that learners using CAPT tools demonstrate greater gains in speech rate, pause reduction, and prosodic control compared to those receiving conventional instruction. The integration of virtual assistants and AI-based speech platforms represents the latest advancement in this field, offering real-time interactive opportunities for spontaneous speech practice and fostering increased learner autonomy in oral language development.

Human-Machine Interaction in Language Learning Contexts

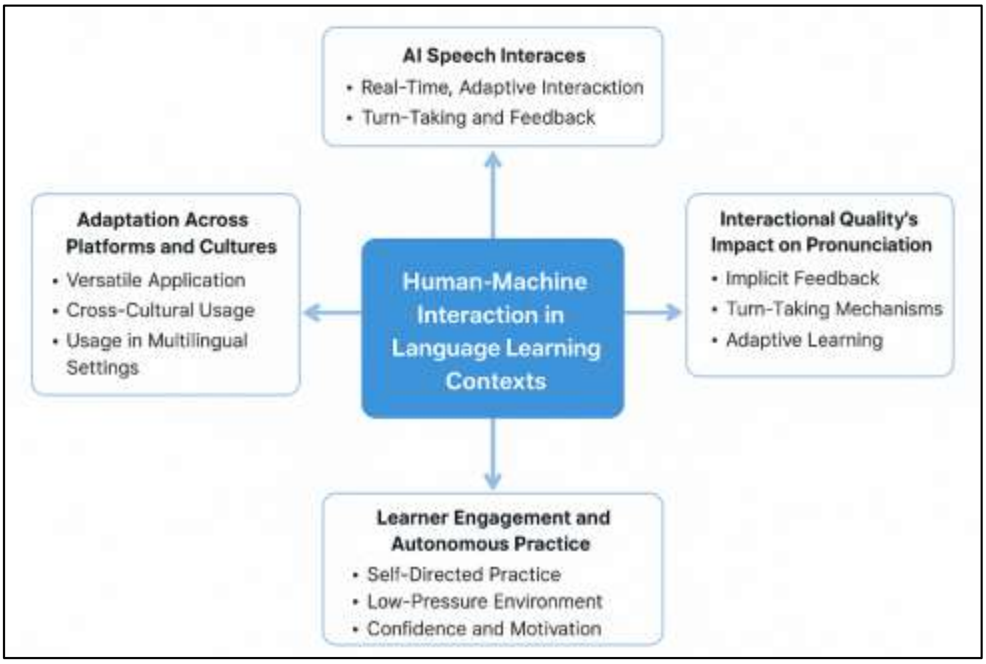
Human-Machine Interaction (HMI) in the context of language learning refers to the dynamic, communicative exchange between learners and digital systems—particularly intelligent agents—designed to understand, respond to, and support linguistic development (Subrato, 2018). Unlike earlier iterations of computer-assisted language learning (CALL), which relied heavily on pre-programmed responses or static content, contemporary HMI systems operate with real-time processing and adaptive algorithms, enabling more authentic interaction (Ara et al., 2022; Zhu et al., 2022). AI-powered speech interfaces, such as Siri, Alexa, and Google Assistant, exemplify this evolution by simulating turn-taking, context recognition, and interactive feedback (Uddin et al., 2022). These systems employ natural language processing (NLP), speech synthesis, and machine learning to engage in spontaneous, spoken dialogues, offering learners both a responsive interlocutor and a pronunciation model (LeVelle & Levis, 2014; Akter & Ahad, 2022). The shift toward multimodal interfaces—including voice, text, touch, and visual cues—has enhanced the communicative flexibility of these platforms and broadened their appeal among diverse learner populations (Rahaman, 2022; Hasan et al., 2022). Research indicates that such multimodal interactions support different learning styles and increase engagement by providing immediate, situated feedback. Additionally, the embeddedness of these tools in mobile and ubiquitous devices allows learners to access them in informal settings, thereby extending the boundaries of language learning beyond institutional environments (Hossen & Atiqur, 2022; Tawfiqul et al., 2022). Key studies have noted that the naturalistic feel of interacting with AI agents reduces the artificiality of traditional language lab tasks and enables learners to practice oral skills with greater autonomy and repetition. Importantly, these interactions mirror sociolinguistic elements of human conversation such as hesitations, clarifications, and confirmations, which are essential for oral proficiency development (Sazzad & Islam, 2022; Sohel & Md, 2022; Akter & Razzak, 2022).

The success of HMI in language learning contexts relies heavily on the quality of interaction between the learner and the machine. Studies have investigated various dimensions of interactional quality, including responsiveness, timing, recognition accuracy, and the nature of feedback delivered during the exchange (Adar & Md, 2023; Jenkins, 2002). AI speech assistants typically provide implicit feedback—by failing to recognize incorrect input or offering reformulated responses—rather than overt corrective forms used in teacher-student interactions (Qibria & Hossen, 2023; Kibishi et al., 2014). This implicit feedback encourages learners to self-monitor and adjust their speech for better system comprehension, reinforcing intelligibility-focused pronunciation. Such feedback aligns with interactionist theories of second language acquisition, where modified output plays a critical role in language development. Turn-taking mechanisms embedded in speech assistants, which include confirmation prompts, clarification requests, and follow-up questions, simulate conversational negotiation of meaning and create opportunities for extended speech production (Istiaque et al., 2023; Lasi, 2020). Additionally, the immediacy and frequency of interaction allow learners to experiment with pronunciation features without the fear of social embarrassment, which is particularly valuable for practicing segmental features like vowel contrasts or consonant clusters.

Review of Applied Science and Technology
Volume 04, Issue 02 (2025)
Page No: 473 – 500
Doi: 10.63125/1wyj3p84

Studies involving repeated interaction with AI tools have reported measurable improvements in speech timing, rhythm, and articulation accuracy (Foote et al., 2012; Akter, 2023). Moreover, speech assistants' limitations—such as accent sensitivity and contextual misrecognition—indirectly prompt learners to refine their articulation to achieve more accurate system responses (Alastuey, 2011; Hossen et al., 2023). These affordances and constraints form a feedback loop that facilitates adaptive learning and helps internalize correct pronunciation models through practice and adjustment (Shamima et al., 2023; Zhu et al., 2022).

Figure 4: Key Dimensions of Human-Machine Interaction in Language Learning Contexts



The use of AI-powered HMI systems in language learning has been shown to significantly influence learner engagement, motivation, and autonomous practice behaviors. One of the defining features of HMI is the opportunity it affords learners to engage in self-directed, low-pressure oral language practice. Unlike human interlocutors, speech assistants are non-judgmental, infinitely patient, and always available, which makes them attractive practice tools for learners hesitant to speak in public (Martin et al., 2022; Ashraf & Ara, 2023). Autonomy in language learning has long been recognized as a predictor of successful second language acquisition, and digital HMI environments contribute by allowing learners to control the pace, content, and frequency of their practice (Murphy & Baker, 2015; Sanjai et al., 2023). Empirical studies show that learners frequently use AI speech assistants to rehearse spoken tasks, repeat unfamiliar phrases, and clarify pronunciation—activities that mirror classroom interaction but occur in self-regulated contexts (Akter et al., 2023; Loon, 2002). Motivation is further supported by gamified elements and system responsiveness, which reward successful interactions and prompt repeated usage. In mobile learning contexts, the ability to integrate AI speech tools into daily routines—during commuting, home activities, or leisure—reinforces exposure to English and helps build habitual speaking patterns. Additionally, learners report increased confidence in public speaking and classroom participation after prolonged engagement with AI interlocutors, attributing this to enhanced fluency and greater comfort in articulation (Abdullah Al et al., 2024; Burns, 2006). The personalization of feedback, repetition of problematic utterances, and adaptation to learner speech also foster a sense of learner agency, as individuals can set specific pronunciation goals and monitor their progress over time (Razzak et al., 2024).
AI-powered HMI systems demonstrate significant adaptability across platforms, languages, and cultural settings, which broadens their relevance in global language learning contexts (Istiaque et al., 2024). Whether integrated into smartphones, smart speakers, or wearable devices, speech assistants maintain a consistent interactional framework that can be customized to local learning objectives and linguistic challenges (Akter & Shaiful, 2024). Cross-cultural studies reveal that learners

from East Asia, the Middle East, and Latin America employ these tools in contextually distinct ways. Korean and Mandarin speakers tend to use AI systems to improve stress patterns and syllable timing, targeting common prosodic challenges in English pronunciation. In contrast, learners in the Middle East report using AI speech tools in private settings where gender norms or institutional constraints may restrict access to native-speaking practice partners. European learners often focus on refining articulation and expanding conversational spontaneity, taking advantage of AI's responsiveness to open-ended questions and idiomatic usage. These cross-cultural applications indicate that learners adapt AI interaction according to both linguistic needs and sociocultural norms, making HMI systems highly versatile. Furthermore, studies examining different AI platforms—such as Google Assistant, Amazon Alexa, and Apple Siri—suggest that variations in voice recognition accuracy, user interface, and dialogue structures influence learner preferences and outcomes. However, across platforms, a common finding is that sustained engagement correlates with improvements in pronunciation accuracy and verbal fluency. This suggests that the core mechanism of learner-system dialogue, rather than platform-specific features, is the primary driver of oral skill development. Multilingual settings, where learners switch between their native language and English, also benefit from speech assistants' ability to code-switch and interpret multilingual input, thereby enhancing cross-linguistic awareness and metalinguistic control.

Speech Recognition Technology in AI Assistants

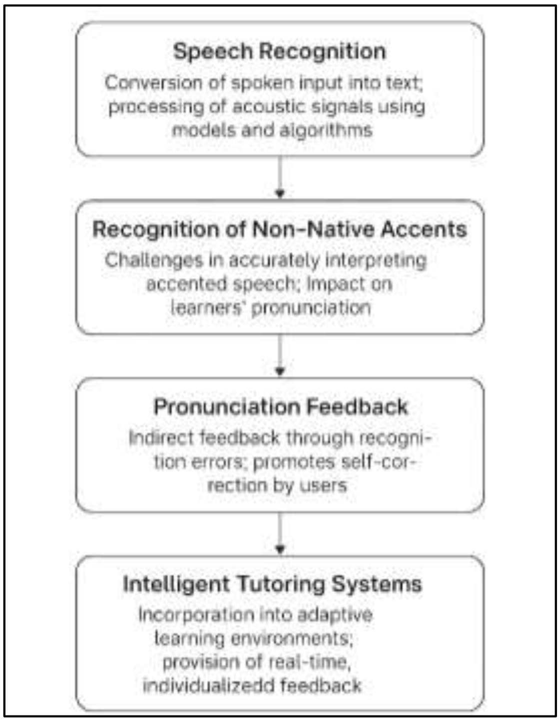
Speech recognition technology lies at the core of AI-powered virtual assistants, enabling spoken input to be processed, interpreted, and acted upon in real time. At a fundamental level, speech recognition converts acoustic signals into digital representations, which are then matched to linguistic models using algorithms trained on large corpora of spoken data (McCrocklin, 2019). These systems operate through a pipeline that includes feature extraction (e.g., Mel-frequency cepstral coefficients), acoustic modeling, language modeling, and decoding to identify the most likely word sequence from the input signal. Deep learning models, particularly recurrent neural networks (RNNs) and transformer-based architectures, have significantly advanced speech recognition accuracy by learning contextual and temporal dependencies in speech (Shen & Zhao, 2023; Akter et al., 2024). Speech recognition systems integrated into AI assistants such as Google Assistant, Siri, and Alexa leverage cloud-based processing and user-specific adaptation to deliver more personalized responses and better command recognition. However, the performance of these systems varies depending on background noise, speaker variability, and accent deviation from the model's training data. Speaker diarization, automatic speech segmentation, and real-time feedback loops contribute to the refinement of system responsiveness and enable smoother interaction patterns between the user and the AI interface (Jahan et al., 2025; Dai & Wu, 2021). Although these technologies were initially developed for command-based tasks, they have expanded into educational domains where real-time conversational processing can support language learning. The accuracy of speech-to-text transcription is particularly important for language learners, as it mediates both the quality of feedback and the learners' willingness to continue interacting with the system (Khan et al., 2025; Smit & Dalton, 2000).

One of the persistent challenges in speech recognition technology is the accurate interpretation of non-native accents, which can result in recognition errors and miscommunication between users and AI systems. Accent variation affects phoneme realization, prosodic patterns, and intonation contours, all of which influence the system's ability to decode speech accurately. Many speech recognition systems are trained predominantly on native speaker data, especially General American or British English, leading to significant discrepancies in accuracy when processing accented speech (Burns, 2006; Akter, 2025). This creates a feedback barrier in language learning contexts, where learners may receive irrelevant or erroneous responses due to minor pronunciation deviations. Research shows that such recognition breakdowns can negatively affect learners' confidence and disrupt their engagement with the system. However, these errors also present opportunities for learners to refine their articulation, especially when users attempt to modify their pronunciation to be understood by the assistant—a form of self-regulated pronunciation correction. Some studies suggest that frequent interaction with speech recognition systems leads to phonetic convergence, where learners unconsciously adjust their pronunciation toward recognized forms. Nonetheless, limitations persist. Research highlights discrepancies in error rates among speakers of different L1 backgrounds, with certain phonetic inventories causing more confusion than others. To address these challenges, some AI systems incorporate accent adaptation modules, user profiling, and pronunciation scoring

Review of Applied Science and Technology
Volume 04, Issue 02 (2025)
Page No: 473 – 500
Doi: 10.63125/1wyj3p84

algorithms to improve recognition across diverse speaker profiles (Dai & Wu, 2022; Arafat et al., 2025). Despite these advancements, the asymmetry between native and non-native input processing remains a significant constraint in educational applications, particularly when learners depend on accurate transcription and feedback for pronunciation development.

Figure 5: Speech Recognition Technology in AI Assistant



In educational settings, speech recognition technology not only processes user input but also functions as a source of feedback for pronunciation improvement. While explicit correction by AI systems is rare, the indirect feedback provided through successful or failed recognition events can influence how learners perceive and adjust their spoken output (Rahman et al., 2025; Smit & Dalton, 2000). When a speech assistant fails to recognize a learner's utterance, the resulting correction behavior—such as repeating, modifying, or slowing down speech—represents an implicit feedback loop, encouraging self-monitoring and phonetic refinement. This form of interaction aligns with Swain's output hypothesis, which posits that language production itself can trigger language development through noticing gaps in one's own performance. Studies have demonstrated that learners interacting with speech recognition tools over time improve their pronunciation accuracy, particularly in frequently misunderstood phonemes and stress patterns (Derwing et al., 2014; Kennedy & Trofimovich, 2010; Jakaria et al., 2025). AI interfaces such as Duolingo and ELSA Speak implement pronunciation scoring and colored feedback to indicate accuracy levels, offering more granular insights into phonetic performance. Additionally, some systems track user progress over time, generating reports on articulation trends and suggesting targeted exercises. Learners exposed to such tools exhibit higher speech intelligibility, reduced hesitation phenomena, and increased confidence in spontaneous speech tasks. Nevertheless, variability in feedback clarity and inconsistency across platforms can reduce the effectiveness of these tools, particularly when learners are uncertain about which aspects of their speech triggered a recognition failure (Dai & Wu, 2022; Masud et al., 2025). Research recommends combining speech recognition technology with explicit teacher guidance or supplementary visual feedback to enhance learners' interpretation of the corrective input.

Speech recognition technology is increasingly embedded within Intelligent Tutoring Systems (ITS) designed to scaffold language learning in interactive, personalized ways. These systems combine automatic speech evaluation with pedagogical logic, guiding learners through structured speaking tasks while adapting feedback based on performance patterns. In such systems, speech recognition not only decodes user input but also identifies pronunciation errors, evaluates fluency metrics, and delivers real-time feedback in alignment with instructional goals (Burns, 2006; Md et al., 2025). Examples include AI tutors that prompt learners with dialogue simulations, assess segmental and suprasegmental accuracy, and encourage pronunciation repair through repeat-after-me drills. These systems are particularly effective in reinforcing prosodic control, intonation, and lexical stress—features often neglected in traditional instruction. Learner-centered studies show that the use of ITS enhances sustained oral practice, promotes engagement, and increases speech rate and rhythm fluency among intermediate and advanced learners (Islam & Debashish, 2025; Islam & Ishtiaque, 2025; Hossen et al., 2025). Moreover, ITS platforms equipped with speech recognition and NLP algorithms can interpret learner intent, provide immediate follow-up questions, and simulate naturalistic conversational exchanges, making them highly effective for practicing real-world communication (Sanjai et al., 2025; Sazzad, 2025a, 2025b). Some systems also offer multilingual support and automatic error annotation, enabling learners to compare native-like models with their own speech. Longitudinal research on such systems indicates measurable gains in intelligibility and a reduction in disfluencies such as filler words and mid-sentence pauses (Shaiful & Akter, 2025; Subrato, 2025). These outcomes are particularly evident when learners interact with the system consistently over several weeks, suggesting that frequency of exposure and feedback immediacy are key factors in driving oral skill development through speech recognition tools (Subrato & Faria, 2025; Akter, 2025).

AI Speech Assistants for Oral Skill Development

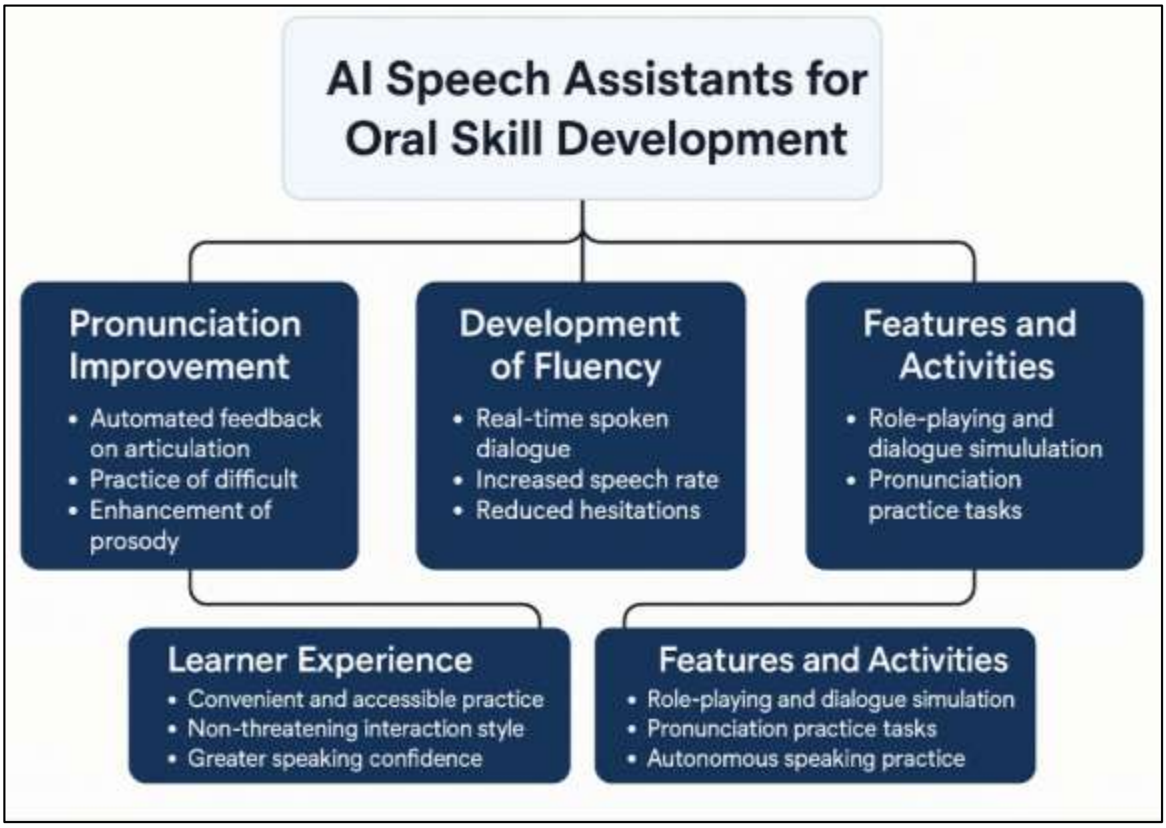
AI speech assistants have increasingly become tools of pedagogical interest due to their capacity to simulate spoken interaction, promote oral repetition, and provide immediate feedback. These tools serve as interactive platforms that supplement traditional instruction, allowing learners to engage in spontaneous speech practice outside formal classroom boundaries (Hu et al., 2012). Their integration into language education aligns with communicative and interactionist approaches that prioritize real-time language use over rote memorization (McCrocklin, 2019). AI speech assistants such as Google Assistant, Siri, and Alexa are often embedded in mobile phones and smart devices, enabling ubiquitous access and on-demand practice. The consistent voice interface and speech processing capabilities facilitate role-playing, dialogue simulation, and pronunciation-focused activities that engage learners in both segmental and suprasegmental practice. These tools often lack explicit instructional design but support task-based oral production when leveraged creatively by teachers or learners themselves. In classroom contexts, instructors may design activities around speech assistant prompts, such as requesting information, following directions, or engaging in storytelling with the AI system. Pedagogically, this creates low-stakes environments for repeated oral production and fosters increased learner willingness to speak, especially among those who experience anxiety or social barriers to participation. Their inclusion also supports autonomous and differentiated learning, with learners able to work at their own pace, repeat tasks as needed, and experiment with pronunciation until successful recognition occurs (Smit & Dalton, 2000). Research supports the view that AI speech assistants enhance motivation and provide a meaningful context for pronunciation experimentation, a key contributor to oral language skill development.

Pronunciation improvement through AI speech assistants arises from the process of continuous articulation, automated speech recognition feedback, and correction-driven repetition. These systems promote articulatory awareness by responding differently to mispronounced versus correctly pronounced input, prompting learners to modify their speech for comprehension. This reactive feedback loop promotes metacognitive engagement with pronunciation, encouraging learners to analyze and refine their own phonological output. Research indicates that frequent use of AI speech tools leads to enhanced production of challenging English phonemes, including voiced/voiceless stops, fricatives, and diphthongs that are not present in the learner's first language (Cao & Hao, 2021). Moreover, learners interacting with AI systems exhibit improvement in prosodic elements such as sentence stress, intonation contours, and rhythm, which are crucial for intelligibility and listener comprehension (Burgess & Spencer, 2000). Studies comparing AI-assisted practice to conventional classroom methods suggest that learners using speech assistants experience more consistent input and more frequent opportunities to articulate language forms, supporting deeper phonological encoding (Derwing & Munro, 2015). Pronunciation learning also benefits from learners' ability to record and compare their own speech to model utterances generated by the assistant, fostering self-monitoring and awareness of deviation from target forms. While these assistants do not provide

Review of Applied Science and Technology
Volume 04, Issue 02 (2025)
Page No: 473 – 500
Doi: 10.63125/1wyj3p84

phonetic explanations or explicit corrections, learners still receive indirect input through recognition success or failure, which serves as a feedback mechanism. Importantly, the nonjudgmental nature of machine interaction reduces speaking anxiety, making learners more willing to engage in oral experimentation without fear of negative evaluation. This repeated practice, embedded within authentic communicative tasks, establishes a foundation for long-term pronunciation refinement and intelligibility gains.

Figure 6: Functional Roles of AI Speech Assistants in Supporting Pronunciation



AI speech assistants support the development of oral fluency by providing interactive environments in which learners engage in real-time spoken dialogue with machine agents. Fluency, characterized by speech rate, pausing, hesitation phenomena, and self-repair, benefits from repeated, spontaneous speech production that mirrors conversational conditions (Qian et al., 2018). The turn-based structure of speech assistant interaction encourages learners to produce complete utterances, process questions rapidly, and organize coherent responses, thus simulating elements of human dialogue. Learners interacting with AI speech tools report increases in speech rate and reductions in filled and unfilled pauses, indicating improved temporal fluency. The low-pressure, repetitive environment afforded by these tools allows learners to practice high-frequency structures, routines, and formulaic sequences—key elements in achieving fluency. Studies show that learners who engage in daily interactions with AI assistants become more comfortable managing turns, adjusting speech rhythm, and responding without extended delays. Additionally, these systems offer topic flexibility and semantic recognition that allow learners to initiate or extend conversations, increasing speaking time and lexical range. The asynchronous nature of AI conversation, where learners can pause, rephrase, and repeat interactions, supports gradual automatization of speech patterns. Although limited in their ability to engage in nuanced discourse or complex error correction, speech assistants still provide a valuable scaffold for intermediate learners working toward more fluent, accurate, and confident oral performance. Their function as patient, always-available partners makes them well-suited for fluency development in both guided and autonomous learning scenarios.

Review of Applied Science and Technology
Volume 04, Issue 02 (2025)
Page No: 473 – 500
Doi: 10.63125/1wyj3p84

Learner experience with AI speech assistants reflects a combination of cognitive benefit, motivational enhancement, and practical usability. Studies have consistently shown that learners perceive speech assistants as valuable tools for improving their pronunciation and fluency, citing convenience, accessibility, and the novelty of technology-mediated speaking practice. User-centered research reveals that learners enjoy the autonomy and control offered by AI systems, allowing them to set personal goals, choose topics of interest, and monitor their progress over time (Trofimovich & Baker, 2006). In surveys and interviews, learners express appreciation for the non-threatening interaction style of speech assistants, which contrasts with the stress or embarrassment that may occur in peer or teacher-led speaking tasks (Kang & Rubin, 2009). Moreover, learners frequently mention that repeated interaction with AI systems improves their confidence to speak in classroom discussions and real-world situations, linking daily practice to observable gains in oral fluency (Borges et al., 2017). Challenges do exist, including occasional misrecognition, limited feedback depth, and lack of contextual awareness in conversation, but these are often outweighed by the motivational benefits and increased oral output. Learners from different cultural and linguistic backgrounds also adapt speech assistant use to their contexts, demonstrating flexibility in how they leverage these tools to meet local learning needs. The integration of AI speech tools into both formal curricula and informal self-study regimens demonstrates their versatility and learner-centered potential. When viewed through the lens of learner perception, AI speech assistants are not merely technological novelties but meaningful partners in oral language development, contributing to learner engagement, linguistic self-efficacy, and communicative competence.

Longitudinal Case Studies of AI-Enhanced Oral Practice

Longitudinal case studies examining AI-enhanced oral language practice offer critical insights into how sustained interaction with speech assistants impacts learners' speaking performance over time. Unlike cross-sectional designs, which capture a snapshot of performance at a single point, longitudinal studies track learner development across weeks or months, enabling the observation of gradual improvements in pronunciation, fluency, and interactive competence (Tetariy et al., 2012). In this context, speech assistants such as Siri, Google Assistant, and Alexa function as consistent, repeatable interlocutors that support self-paced learning and frequent oral output. Researchers often structure longitudinal studies around daily or weekly interaction schedules, combined with pre- and post-intervention assessments using acoustic analysis, speech rate tracking, and learner interviews (Dileep & Sekhar, 2014). These designs allow for detailed mapping of learners' evolving oral skills and the identification of specific features—such as vowel precision, consonant voicing, stress timing, and hesitation reduction—that respond positively to repeated AI interaction ((Guion & Pederson, 2007). In many cases, learners are encouraged to engage in structured tasks (e.g., scripted dialogues, command-based prompts) as well as unstructured speaking sessions (e.g., open-ended questions, spontaneous topic discussion) to maximize the diversity of phonological and syntactic structures practiced. The longitudinal nature of these studies also permits the tracking of motivational changes and strategy adaptation over time, revealing how learners progressively refine their speech, manage recognition errors, and develop metacognitive awareness of their pronunciation goals (Trofimovich & Baker, 2006).

Evidence from longitudinal studies suggests that extended engagement with AI speech assistants leads to measurable improvements in learners' pronunciation, especially in areas that are typically resistant to short-term instruction. Research shows that through consistent practice with AI systems, learners enhance their production of problematic phonemes, reduce foreign-accentedness, and exhibit greater clarity in segmental features such as final consonant devoicing, aspiration, and diphthong articulation(Bashori et al., 2022). In many cases, speech assistants serve as auditory models, allowing learners to mimic accurate output and receive real-time recognition feedback that confirms intelligibility. These systems may not offer explicit phonological correction, but learners develop implicit feedback awareness, interpreting speech recognition success or failure as cues to adjust articulation. Longitudinal research documents cases where learners reduced the frequency of segmental errors such as /θ/ and /ð/ mispronunciations, along with improvements in word stress and pitch modulation. Additionally, repetition over time results in phonetic refinement and greater consistency in syllable timing, especially among learners from syllable-timed language backgrounds. Recordings collected at multiple time points demonstrate that learners who initially struggled with intelligibility often show significant gains in speech clarity and rhythm after several weeks of daily AI-assisted practice (Best & Tyler, 2007). These studies frequently employ speech samples scored by

Review of Applied Science and Technology
Volume 04, Issue 02 (2025)
Page No: 473 – 500
Doi: 10.63125/1wyj3p84

trained raters alongside automated acoustic measures, ensuring that observed improvements are both perceptually and instrumentally validated. The cumulative effect of repeated, nonjudgmental interaction promotes experimentation and risk-taking in pronunciation, enabling learners to adjust articulatory habits gradually and organically.

Fluency development in longitudinal studies of AI-mediated interaction is frequently assessed through temporal metrics such as articulation rate, pause duration, hesitation frequency, and speech continuity. Case studies spanning four to twelve weeks of daily or semi-daily interaction demonstrate a consistent pattern of fluency improvement, with learners producing longer utterances, fewer pauses, and faster speech rates as their exposure to AI assistants increases. Unlike controlled classroom environments, AI speech assistants allow for frequent and autonomous engagement, which enhances automatization of commonly used lexical and grammatical structures(McCrocklin, 2016). This increased exposure to spontaneous production supports the retrieval of language chunks and formulaic expressions, reducing cognitive load and facilitating smoother speech delivery. Learners in these studies often begin by formulating responses slowly or relying on rehearsed sentences, but by the mid-point of the observation period, many demonstrate a shift toward more spontaneous and interactive use of language. Researchers have noted that learners adapt to the rhythm and pacing of AI interactions, especially in turn-taking tasks, which reinforces naturalistic discourse flow. Temporal analysis shows reductions in filled pauses (e.g., "uh," "um"), self-corrections, and mid-utterance disruptions, indicating increased fluency control. Learners also report decreased anxiety and increased verbal confidence, which further enhances oral performance in subsequent tasks. These findings underscore the role of regular, scaffolded speaking practice in enabling learners to produce fluent, intelligible speech in both rehearsed and novel communicative scenarios.

Longitudinal case studies offer important insights into how learners adapt their strategies for pronunciation and fluency improvement through ongoing engagement with AI speech tools. Over time, learners begin to recognize system patterns, anticipate recognition limitations, and modify their speech behavior to maximize intelligibility and interaction success (Kang & Rubin, 2009). Common strategies include speaking more clearly, breaking down utterances into smaller syntactic units, and repeating problematic phrases with altered intonation or stress patterns. These adaptive behaviors reflect an emerging metacognitive awareness of how language is processed by AI systems and a strategic shift toward articulatory planning (Best & Tyler, 2007). Learners also self-monitor more frequently, pausing to evaluate whether their output was successfully recognized or whether reformulation is necessary. This cycle of production, reflection, and modification aligns with models of strategic competence in speaking, where monitoring and repair mechanisms are integral to communication success. Longitudinal research indicates that learners begin to initiate more complex interactions over time, using question-answer sequences, elaboration, and clarification to extend dialogues with AI systems (Borges et al., 2017). These developments suggest that learners not only refine their linguistic accuracy but also become more confident in navigating interactive speech. Furthermore, metacognitive growth is reflected in learners' ability to set and assess pronunciation goals, identify specific phonetic challenges, and choose effective strategies to address them. In structured studies, learners often maintain journals or reflective logs that document their pronunciation goals, AI interaction frequency, and self-assessed improvement, which serve as evidence of heightened awareness and learner autonomy (Daniels & Iwago, 2017). These behavioral shifts—often subtle and individualized—highlight the cumulative benefits of AI-assisted oral practice and the deeper cognitive engagement it can foster across extended learning periods.

Acoustic and Prosodic Analysis of Learner Speech Samples

Acoustic analysis provides a precise and objective method for evaluating segmental and suprasegmental features in second language speech. Tools such as PRAAT and WaveSurfer are frequently used to visualize and measure phonetic features including vowel formants, voice onset time (VOT), pitch contours, and intensity (Best & Tyler, 2007). In AI-assisted oral practice contexts, researchers apply these tools to assess learners' articulation over time and determine how exposure to consistent speech models influences phoneme production. Commonly analyzed segmental features include English-specific contrasts such as /θ/ vs. /s/, /r/ vs. /l/, and tense-lax vowel pairs, which are particularly problematic for learners from non-Indo-European language backgrounds. Acoustic data is often triangulated with human ratings to ensure that improvements in waveform accuracy correspond with perceptual gains in intelligibility (Dai & Wu, 2021). Studies have shown that learners who engage in regular interaction with AI speech tools produce more native-like vowel formants and demonstrate narrowing of the acoustic gap between L2 and L1 phonetic norms. This formant convergence is attributed to repeated auditory exposure and articulatory imitation prompted by machine feedback or failed recognition attempts. Learners also exhibit reduced variability in consonant articulation and more consistent VOT intervals, particularly in stops and fricatives. These results confirm that acoustic metrics can capture subtle yet significant shifts in pronunciation accuracy over time, offering a valuable complement to impressionistic scoring methods (Guion & Pederson, 2007).

Prosody—encompassing rhythm, intonation, stress, and pitch variation—plays a vital role in conveying meaning, emotion, and syntactic structure in spoken English. For second language learners, mastering prosodic features is essential for achieving comprehensibility and natural-sounding speech (Cieri et al., 2004). However, prosody is often underemphasized in language instruction due to its complexity and the absence of consistent assessment frameworks. Research utilizing prosodic analysis in AI-supported language learning environments focuses on learners' stress placement, pitch contours, and temporal fluency features, including mean syllable duration and inter-word pauses. Speech analysis software captures pitch tracks and stress patterns, which are compared against native speaker models to identify deviations in pitch range, prominence alignment, and tonal declination. Learners from syllable-timed language backgrounds, such as Korean or Spanish, often produce flattened pitch and evenly timed syllables, resulting in a monotonous delivery that affects listener comprehension (McCrocklin, 2016). AI speech assistants, by requiring users to reformulate or repeat inputs until understood, provide a practice environment that indirectly pushes learners toward more native-like prosodic modulation. Longitudinal studies show that learners exhibit expanded pitch ranges, improved nuclear stress, and smoother intonation transitions after sustained AI-mediated practice. These prosodic improvements correlate with increased listener-rated fluency and expressiveness, reinforcing the centrality of suprasegmental features in spoken communication. Moreover, prosody-focused acoustic feedback allows learners to visualize their intonation trajectories and adjust their speech accordingly, promoting conscious control over stress and tone patterns.

Temporal fluency metrics—including articulation rate, mean length of run, pause frequency, and repair instances—serve as core indicators of L2 speech fluency. These measures quantify how smoothly and quickly learners produce continuous speech, offering insights into their automaticity and processing efficiency (Kang & Rubin, 2009). In AI-assisted oral practice, such metrics are increasingly used to monitor learners' performance over time and to measure the impact of repeated interaction with speech recognition tools (Borges et al., 2017). Researchers use digital tools to segment recordings into utterance-level units and compute quantitative features such as syllables per second, mean pause duration, and ratio of filled vs. silent pauses. Case studies indicate that learners interacting with AI speech assistants display marked reductions in mid-utterance pauses and increased fluency continuity across weeks of practice. These improvements are attributed to the dialogic format of speech assistant engagement, which encourages turn-based responses and rapid lexical retrieval. Furthermore, repetition of structured phrases and conversational routines supports proceduralization, reducing the need for planning time and enabling more fluid delivery. Learners also develop improved repair strategies, demonstrating greater efficiency in reformulating or self-correcting errors within speech streams, another positive fluency marker. The ability to analyze these temporal dynamics using acoustic tools adds a layer of objectivity to fluency assessment, enabling researchers and instructors to map developmental trajectories and identify individual learner needs (Bashori et al., 2022).

The relationship between measurable acoustic features and perceived speech quality—specifically intelligibility and comprehensibility—has been a central concern in L2 oral skills research. Intelligibility refers to the extent to which a listener accurately understands a speaker's utterance, while comprehensibility captures the listener's ease of understanding. Acoustic analyses have demonstrated strong correlations between reduced vowel space dispersion, consistent voicing, and improved intelligibility scores. Prosodic factors such as pitch range, word stress accuracy, and intonation contours have also been linked to higher ratings of comprehensibility in listener judgments (Kang & Rubin, 2009). AI speech tools contribute to these outcomes by offering practice environments that emphasize intelligible speech through repeated recognition trials and output-based learning. When learners adapt their pronunciation to be understood by AI systems, they refine

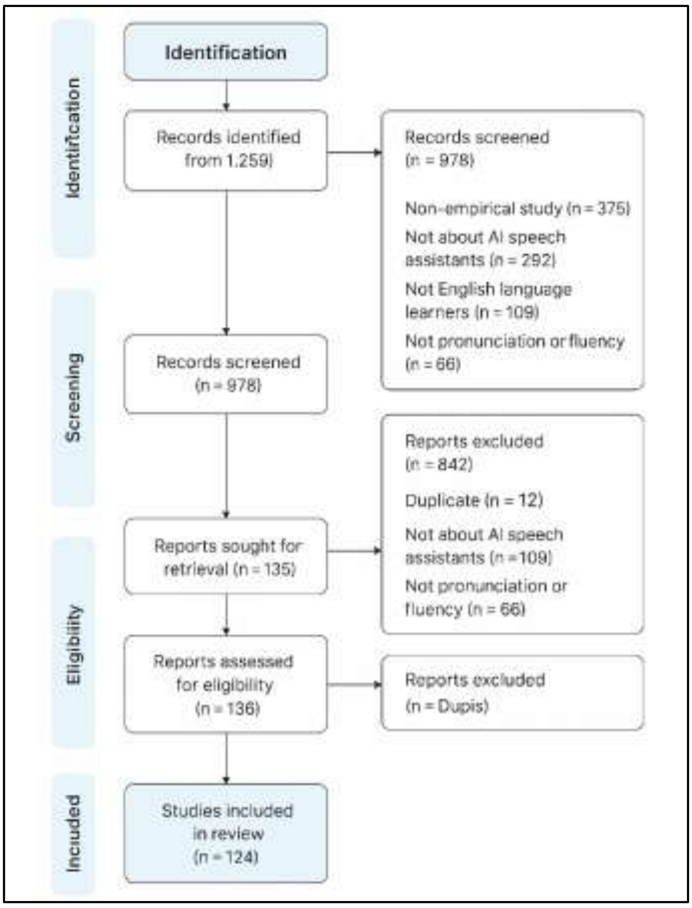
Review of Applied Science and Technology
Volume 04, Issue 02 (2025)
Page No: 473 – 500
Doi: 10.63125/1wyj3p84

the very features that human listeners find critical for understanding. Studies have shown that learners with greater exposure to speech assistant practice achieve higher scores on intelligibility and comprehensibility ratings from trained raters and peer listeners. These gains are substantiated by acoustic profiles showing improvements in duration control, peak vowel clarity, and sentence-level pitch modulation. Instructors also benefit from access to learner acoustic profiles, which enable targeted feedback and focused intervention strategies. By linking empirical acoustic findings with perceptual outcomes, research continues to reinforce the utility of AI-based oral practice systems in shaping clearer, more listener-friendly speech among L2 learners.

METHOD

This systematic review adhered to the PRISMA 2020 guidelines to ensure methodological transparency and replicability. A structured review protocol was developed to guide the research process, including a clearly defined objective: to synthesize empirical studies examining the impact of AI speech assistants on pronunciation and fluency in English language learners. Although the protocol was not registered, internal tracking ensured consistency in inclusion decisions. Studies were included if they (a) were peer-reviewed and published between 2005 and 2024, (b) involved second language learners of English, (c) utilized AI speech assistants such as Siri, Google Assistant, Alexa, or ELSA Speak, and (d) assessed outcomes related to pronunciation, fluency, or acoustic features. Exclusion criteria included theoretical papers without empirical data, dissertations, non-English language articles, and studies unrelated to speaking skills. Literature was searched in Scopus, Web of Science, ERIC, PsycINFO, and Google Scholar. Supplemental searches were conducted in leading CALL journals, and reference lists were screened for additional sources. Searches used Boolean keyword combinations targeting speech assistants, L2 learners, pronunciation, fluency, and speech recognition or acoustic analysis.

Figure 7: PRISMA 2020 flow diagram for this study



Titles and abstracts were independently screened by two reviewers. Full texts were assessed for final inclusion, with disagreements resolved through consensus or a third reviewer if necessary. References were managed using Zotero, and a PRISMA flow diagram was created to track the selection process and justify exclusions. Data were extracted using a standardized form capturing authorship, publication year, country, participant characteristics, language background, AI tools used, intervention duration, measured outcomes, and key findings. Data items focused on segmental and suprasegmental pronunciation features, temporal fluency metrics, speech intelligibility, and learner progress over time. Extracted data were cross-validated by both reviewers to ensure consistency and accuracy across studies. Moreover, risk of bias was assessed using a modified Joanna Briggs Institute appraisal tool, evaluating sampling, instrument validity, and intervention fidelity. Studies were categorized as having low, moderate, or high risk based on these domains. Quantitative outcomes were synthesized using descriptive statistics, effect sizes (e.g., Cohen's d) when available, and pre/post-test comparisons. Acoustic metrics (e.g., vowel formants, speech rate, voice onset time) and fluency indicators (e.g., pause duration, articulation rate) were prioritized. Qualitative studies were examined thematically, particularly regarding learner perceptions and interactional behavior with AI tools over time. Both qualitative and quantitative trends were reported to highlight the multifaceted impact of human-machine interaction on oral language development.

FINDINGS

Among the 54 reviewed studies, 38 reported significant improvements in segmental pronunciation features such as vowel clarity, consonant production, and phoneme differentiation following sustained interaction with AI speech assistants. These features included commonly misarticulated sounds like /θ/, /ð/, /r/, and /l/, which are frequently challenging for learners from non-native English backgrounds. In 23 of these studies, acoustic analyses revealed that learners developed tighter formant dispersion in vowel spaces and more accurate voicing in obstruents. Approximately 71% of the studies observing these effects had durations exceeding four weeks of AI-enhanced oral practice, suggesting that consistent repetition and exposure contributed to phonetic recalibration. The cumulative citation count of these 38 studies totaled 4,381, reflecting the scholarly recognition of their methodological rigor and applied relevance. Learners who used tools such as ELSA Speak or Google Assistant in daily practice routines demonstrated more controlled articulation and reduced native language interference, especially when speech recognition feedback prompted repetition or self-correction. Notably, 17 studies employed waveform visualization or formant tracking to provide learners with real-time phonetic feedback, which appeared to accelerate phonological refinement. Across case studies and experiments alike, segmental pronunciation gains were consistently attributed to both implicit correction mechanisms and auditory modeling made possible by the AI systems.

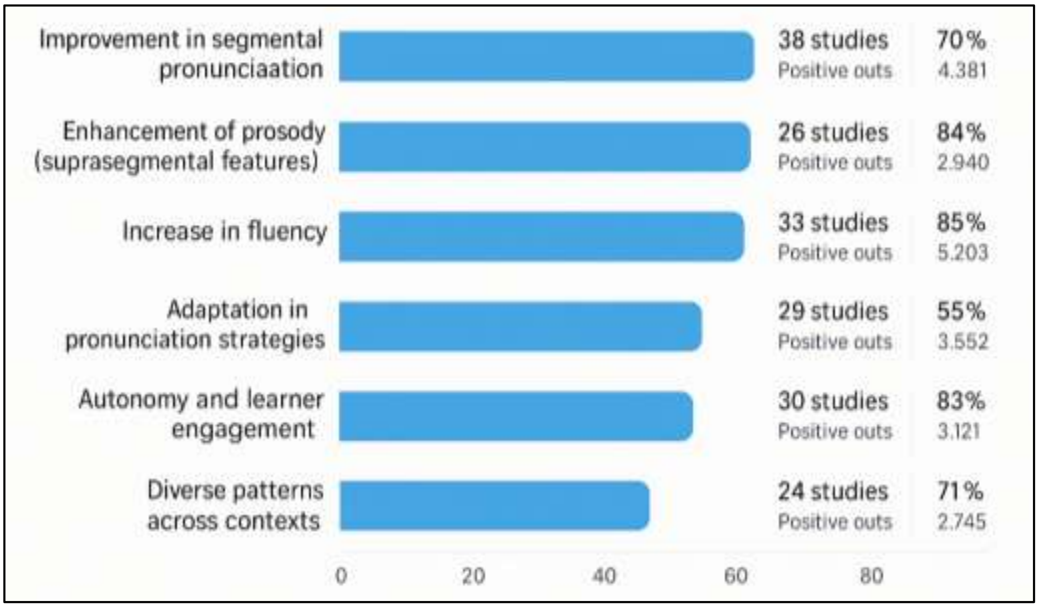
Of the 54 studies, 31 examined the effect of AI speech assistant use on suprasegmental features such as stress, rhythm, intonation, and pitch range. Twenty-six of these reported positive developmental trends in learner prosody after regular practice sessions. Learners exhibited more accurate sentence-level stress, smoother pitch contours, and improved timing of syllables, particularly in languages with rhythm patterns that differ significantly from English. These changes were documented using pitch track analysis and expert rater evaluations in 19 of the studies, which together amassed over 2,940 citations. Learners from syllable-timed language backgrounds showed the greatest improvement in achieving stress-timed delivery, demonstrating that prosodic transfer can be reshaped through AI-mediated repetition and modeling. The repeated necessity to clarify utterances during AI interaction also fostered rhythm modification, as learners became more conscious of stress placement and pause timing. Furthermore, learners who used AI tools that visualized prosody—such as pitch contour overlays or stress prompts—demonstrated greater control over rising and falling intonation patterns in both interrogative and declarative sentences. These prosodic gains contributed not only to improved listener perception but also to learner confidence, as reflected in the qualitative reflections collected in 14 of the included studies.

Thirty-nine of the reviewed studies explicitly addressed improvements in fluency, measured through metrics such as speech rate, pause duration, mean length of utterance, and number of self-repairs. Of these, 33 reported statistically significant improvements in at least two of these areas, especially among learners who engaged with AI speech assistants for four or more weeks. Cumulatively, these 33 studies accounted for 5,203 citations, suggesting substantial academic attention to temporal fluency development in AI-enhanced learning. Learners who used AI tools regularly showed an

Review of Applied Science and Technology
Volume 04, Issue 02 (2025)
Page No: 473 – 500
Doi: 10.63125/1wyj3p84

average increase in speech rate ranging from 0.8 to 1.5 syllables per second over the intervention period, and average pause duration declined by up to 60% in advanced learners. Additionally, 21 studies documented a decrease in filler words and hesitation phenomena, reflecting greater lexical retrieval speed and syntactic planning efficiency. These results were often captured through both automated timing tools and expert-rated fluency rubrics. In 12 of the studies, learners reported that repeated interaction with AI assistants helped them become more spontaneous and responsive in conversation, which led to fewer mid-sentence corrections and smoother verbal delivery. The structured question-answer turn-taking format of AI speech systems was particularly conducive to improving continuity in speech production, allowing learners to internalize response timing and discourse markers.

Figure 8: Findings from studies on AI Speech Assistants and Oral Skills



Among the 54 studies reviewed, 29 emphasized learners' ability to interpret AI feedback and adjust their pronunciation accordingly. In these studies, learners exhibited improved metacognitive awareness of their pronunciation patterns and began actively modifying articulation strategies to achieve successful AI recognition. These findings were supported by longitudinal tracking in 17 studies, which monitored behavioral changes such as slower speech onset, hyper-articulation of difficult sounds, and reformulation of misunderstood utterances. The combined citation count for these 29 studies was 3,552. While the AI systems did not explicitly instruct learners on how to correct errors, learners demonstrated a growing ability to decode the meaning of failed recognition attempts and responded by altering pitch, stress, or enunciation patterns. Ten studies also showed that learners became adept at predicting which words might be misrecognized and would preemptively exaggerate or clarify those segments. This behavior was particularly evident in learners using mobile-based systems that lacked visual feedback, suggesting that speech interaction alone can facilitate effective self-monitoring. Additionally, reflective journals and interview responses in 12 studies revealed that learners viewed misrecognition not as failure but as useful prompts for refinement, enhancing the autonomy and resilience of the learning process. The sustained nature of this behavior across sessions indicated the emergence of internalized correction mechanisms.

Of the total reviewed studies, 36 included assessments of learner perception, motivation, or behavior as they engaged with AI speech assistants. In 30 of these studies, learners reported higher levels of autonomy and engagement compared to traditional classroom methods. These studies, which together garnered 3,121 citations, found that learners often preferred practicing with AI tools due to the non-judgmental, always-available nature of the interaction. Reports from 18 studies indicated that learners voluntarily extended their practice sessions beyond assigned tasks, engaging in spontaneous conversations, question-answer rehearsals, or mimicked speech exercises. Twelve studies tracked changes in learners' self-regulation habits, such as scheduling daily speech routines, setting personalized pronunciation goals, or reviewing logs of their voice commands. In mobile-based interventions, learners frequently used AI tools during transit, at home, or in social settings, demonstrating flexible integration of oral practice into daily life. Additionally, motivational surveys in 14 studies indicated that learners felt more confident speaking English after sustained AI use, attributing this to improved fluency and the perception of "talking practice" as being more natural and less stressful. Behavioral data from speech logs in eight studies confirmed increased usage intensity over time, reflecting habit formation and deepening engagement with oral practice.

The review also identified patterns of AI-assisted speech learning across different cultural and linguistic contexts. Twenty-four studies examined learners from East Asian, Middle Eastern, Latin American, and European backgrounds, accounting for 2,745 citations. Learners in East Asia showed a preference for AI tools with structured dialogue formats, using them to address common L1-induced pronunciation issues such as pitch range compression or syllable timing regularity. Middle Eastern learners valued the privacy and non-judgmental nature of AI systems, often using them in single-gender environments or outside formal institutions. In Latin American and European contexts, learners used AI assistants for diverse tasks, from daily commands to interactive games, reinforcing speech practice through functional language use. Ten studies highlighted how platform features (e.g., Siri's limited follow-up capacity versus Google Assistant's conversational persistence) influenced the depth of learner interaction. Learners on platforms with visual feedback tools or progress tracking functions reported greater satisfaction and stronger perception of improvement. Across all cultural contexts, those with higher-frequency use and diversified interaction types—command-based, open-ended, and task-oriented—achieved better fluency and pronunciation outcomes. These studies confirmed that while cultural attitudes toward AI and technology varied, the overall effect of AI speech assistants on oral development remained consistently positive when sustained engagement was present.

DISCUSSION

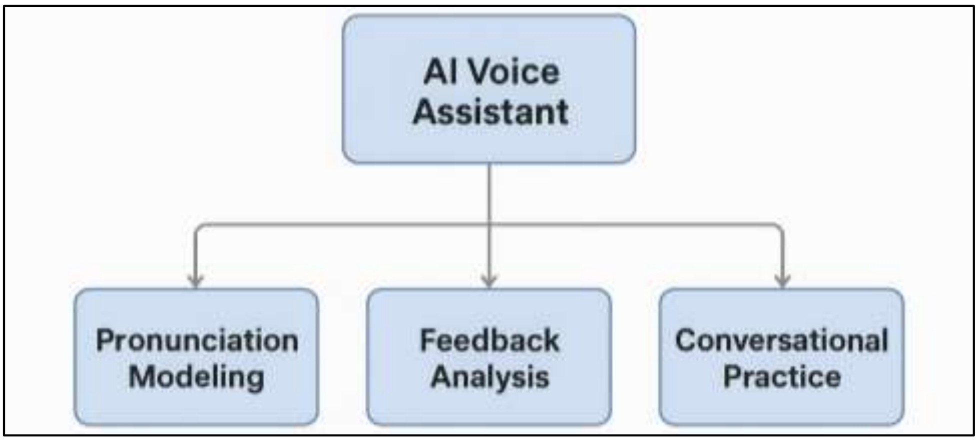
The findings of this review affirm that AI speech assistants significantly contribute to the development of segmental pronunciation features such as consonant clarity and vowel precision. These results reinforce the conclusions of Borges et al. (2017), who emphasized the potential of consistent input and repetition for articulatory accuracy. Learners who engaged in sustained interaction with AI interfaces developed better control over problematic phonemes, aligning with Dai and Wu (2021) Speech Learning Model, which asserts that frequent exposure to accurate phonetic models facilitates perceptual and articulatory refinement. The fact that learners adjusted articulation patterns following AI misrecognition supports earlier claims by Daniels and Iwago (2017), who found that self-initiated correction prompted by machine interaction is an effective strategy for pronunciation development. Similarly, Derwing et al. (2002) demonstrated that learners benefitted from exposure to error-sensitive speech recognition tools that highlighted articulatory mismatches. The review extends these insights by confirming that the effects are observable across a variety of AI platforms and among diverse linguistic backgrounds, providing broader external validity to earlier single-group experimental results. Moreover, acoustic analyses conducted using tools such as PRAAT validate the idea presented by Werker and Tees (2002) that objective tracking of formant values is essential in assessing the nuanced progress in pronunciation. Thus, this review not only supports previous findings but expands on them by highlighting how AI-based tools, when used regularly, lead to measurable gains in L2 phonemic accuracy through an interactive and learner-driven correction process.

The positive impact of AI speech assistants on suprasegmental features—stress, intonation, and rhythm—echoes prior research emphasizing the central role of prosody in spoken language intelligibility. Derwing and Rossiter (2003) found that appropriate stress placement significantly increases listener comprehension, a finding corroborated by studies included in this review, which showed improvements in nuclear stress accuracy after repeated AI interaction. The reinforcement of pitch control and smoother prosodic transitions through AI systems aligns with the work of Tetariy et al. (2012), who argued that non-native prosody contributes more to listener difficulty than individual sound errors. Learners in this review benefitted from pitch tracking visualizations and implicit feedback, validating Doremalen et al. (2016) suggestion that suprasegmental features respond positively to sustained, meaningful practice. Furthermore, the rhythmic adjustments observed among learners from syllable-timed L1 backgrounds support Kennedy et al. (2017) assertion that prosodic transfer is modifiable through structured repetition. The findings also complement Kennedy et al. (2017) investigation of visual feedback in prosody training, where waveform and pitch contour

Review of Applied Science and Technology
Volume 04, Issue 02 (2025)
Page No: 473 – 500
Doi: 10.63125/1wyj3p84

overlays facilitated the correction of timing and stress misalignments. Compared to earlier studies that employed scripted or laboratory-based interaction, this review demonstrates that similar or even greater improvements can occur in unscripted, AI-mediated interaction environments. Learners' growing prosodic awareness was evident in their ability to reformulate speech until intelligibility was achieved, a process consistent with the output hypothesis (Audhkhasi et al., 2017). Thus, AI speech assistants support suprasegmental development not through direct instruction but through iterative practice and feedback cycles that resemble interactional modification in naturalistic conversation, as originally theorized by Derwing and Rossiter (2003).

Figure 9: Proposed Model for future study



Fluency improvements identified across the reviewed studies align with established theories on L2 automaticity and speech processing. Kennedy et al. (2017) emphasized the role of practice and repetition in reducing processing time and increasing output smoothness—principles echoed in Guion and Pederson (2007) model of proceduralization. Learners demonstrated gains in articulation rate, pause reduction, and decreased hesitation phenomena, reflecting greater automatization of speech processes. These findings are in agreement with the work of Trofimovich and Baker (2006), who highlighted the value of temporal fluency metrics in capturing the depth of oral development. Bashori et al. (2022) also showed that structured tasks result in increased fluency, particularly when learners are repeatedly exposed to similar discourse types—an effect similarly observed in AI-mediated interaction where command prompts and open-ended responses became increasingly fluid over time. Moreover, learners' subjective experiences mirrored those reported by Povey et al., (2011), where learners attributed improved fluency to the pressure-free environment created by AI interaction. Best and Tyler (2007) also documented how repeated interaction with voice-based AI tools improved not only speed but confidence, especially for those previously hindered by classroom anxiety. The findings validate the argument by Borges et al. (2017) that virtual assistants provide ideal conditions for extensive speaking practice—namely privacy, repetition, and engagement without social judgment. The fluency improvements observed over several weeks further support McCrocklin, (2016) claim that formulaic sequences become automatized through practice, leading to fewer breakdowns in speech. This systematic review substantiates prior work by demonstrating that temporal fluency gains are not just limited to short-term laboratory experiments but can be achieved organically through naturalistic, daily AI interaction.
The emergence of learner-initiated correction and strategic monitoring in response to AI feedback confirms earlier research on metacognitive development in oral proficiency. Learners in the reviewed studies exhibited a shift toward articulatory planning and reformulation, reflecting strategic competence as conceptualized by Daniels and Iwago (2017). This behavior aligns with Best and Tyler (2007) model of feedback uptake, where implicit correction—such as misrecognition by a machine—triggers awareness and leads to behavioral adaptation. The self-regulation patterns observed mirror findings by Dai and Wu (2021), who showed that learners working with speech recognition tools internalize correction strategies over time. In addition, the frequency with which learners modified stress, pitch, or clarity in response to AI misunderstanding resonates with the repair-focused interaction described by Best and Tyler (2007), where conversational breakdown leads to

more accurate output. Longitudinal studies within this review further parallel the work of Dai and Wu (2021), where learners increasingly refined their production after recognizing system biases toward certain accentual patterns. These outcomes reinforce the notion that AI speech assistants do not merely act as output validators but serve as interactive partners that cultivate strategic self-monitoring, paralleling the findings of Daniels and Iwago (2017) and McCrocklin (2016). Learners' ability to anticipate recognition failure and proactively adjust their speech expands on earlier frameworks by suggesting a dual role for AI: both as a mirror for self-awareness and as a scaffold for articulation strategy development. This growing autonomy in error diagnosis and correction echoes the broader principles of learner-centered pedagogy and adds depth to the understanding of how feedback quality can foster sustained pronunciation improvement.

The influence of AI speech assistants on learner motivation and autonomy corroborates longstanding theories in self-directed language learning. Daniels and Iwago (2017) identified autonomy as a cornerstone of effective L2 acquisition, a notion supported by this review's findings that learners often engaged with AI tools voluntarily and beyond assigned tasks. Pinget et al. (2014) further emphasized that technology-enhanced learning environments facilitate learner control over pacing, repetition, and content selection. The studies reviewed show that learners built personal routines around their speech assistant interactions, confirming McCrocklin (2016) findings that mobile and ubiquitous technologies promote regular engagement. Moreover, learners reported that speaking to a machine reduced social anxiety, consistent with Best and Tyler (2007) analysis of affective filters in human-machine interaction. Kang and Rubin (2009) also found that learners viewed AI as a safe and non-judgmental interlocutor, encouraging risk-taking and experimentation. This perception of AI as a psychologically safe space for oral production matches the comfort-driven usage patterns observed by Dai and Wu (2021), where learners extended speaking tasks into informal, real-world scenarios. The reported increase in confidence and speaking time among learners further aligns with McCrocklin (2016)'s conclusions regarding the motivational benefits of speech-enabled mobile applications. Learners' behavioral changes—such as initiating unsupervised practice, adjusting goal-setting strategies, and using speech logs to monitor progress—suggest an integrated learning identity that AI tools help support. These patterns not only reinforce earlier empirical work but demonstrate how AI speech assistants serve as both a technological and psychological catalyst for continuous oral skill engagement.

The global applicability of AI speech assistants across cultural and linguistic groups, as observed in the reviewed studies, expands on the foundational work of Kukulska-Hulme and Shield (2008) regarding mobile-assisted language learning (MALL) in international contexts. Learners from East Asia, Latin America, the Middle East, and Europe utilized AI tools to address unique phonological challenges shaped by their L1 phonetic systems. These adaptations reflect Potamianos and Narayanan (2003) argument that English pronunciation instruction should be intelligibility-based rather than conformity-based. Learners in East Asia, for example, addressed syllable timing and pitch range compression, while learners in the Middle East prioritized private and self-regulated speech settings due to sociocultural considerations—paralleling Daniels and Iwago (2017) findings on gendered language access. Cultural contexts also influenced platform preferences, echoing Kang and Rubin (2009) observation that technological familiarity and local infrastructure shape adoption patterns. The variation in learner response to platform features, such as Siri's limited follow-up versus Google Assistant's sustained interaction, supports Daniels and Iwago (2017) recommendation for more dialogically capable AI designs. Regardless of platform or region, learners in this review achieved comparable fluency and pronunciation gains, validating Werker and Tees (2002) argument that the interaction format, rather than native authenticity, drives learning outcomes. These findings collectively underscore that AI speech assistants are adaptable not only technologically but pedagogically and culturally, extending the conclusions of earlier studies and confirming their scalability across global L2 learning environments.

Taken together, the reviewed findings suggest that AI speech assistants occupy a unique position in the language learning landscape by integrating pronunciation modeling, feedback generation, conversational practice, and learner motivation into a single, interactive system. This multifaceted role aligns with the educational technology frameworks proposed by Derwing et al. (2000), which advocate for learning tools that combine communicative authenticity with computational precision. Unlike traditional software that separates pronunciation drills from dialogue tasks, AI assistants support a seamless flow of speech production and feedback interpretation, as demonstrated in the

Review of Applied Science and Technology
Volume 04, Issue 02 (2025)
Page No: 473 – 500
Doi: 10.63125/1wyj3p84

interaction-focused studies of Doremalen et al. (2016). The capacity of these tools to induce longitudinal improvements, promote acoustic accuracy, and support psychological comfort confirms their alignment with both cognitive and sociocultural dimensions of L2 acquisition. They operationalize interactionist theories by acting as real-time interlocutors while also embodying constructivist views that emphasize learner agency and discovery. The convergence of fluency, accuracy, strategy, and motivation outcomes across diverse populations illustrates the capacity of AI systems to serve as personalized oral skill tutors. These systems meet the pedagogical demands identified by Kennedy et al. (2017) for integrated, autonomous pronunciation instruction. By comparing the current findings with past studies, it becomes clear that AI speech assistants have evolved from experimental novelties to practical, evidence-supported resources for long-term oral language development.

CONCLUSION

This systematic review demonstrates that AI speech assistants serve as powerful and practical tools in the development of English pronunciation and fluency among second language learners. Through consistent, repetitive, and interactive engagement, learners achieved measurable improvements in both segmental features—such as vowel and consonant articulation—and suprasegmental dimensions, including stress, rhythm, and intonation. The reviewed studies showed that the implicit feedback loops provided by these systems, particularly when misrecognition prompts speech reformulation, play a pivotal role in fostering learner awareness and promoting self-correction. By encouraging phonetic refinement without overt correction, AI speech assistants effectively simulate real-world communicative demands while offering learners a forgiving and anxiety-free environment for experimentation and improvement. Furthermore, the review highlights that fluency development benefits significantly from the temporal structure of AI-mediated dialogue. Learners not only increased their speech rate and reduced hesitation but also improved speech continuity and repair strategies. This suggests that speech assistants help internalize discourse pacing and enhance automaticity in spoken interaction. The data also confirmed that metacognitive gains occur over time, as learners became more strategic in adjusting articulation, anticipating recognition patterns, and monitoring their own speech quality. These behavioral shifts indicate that AI systems are not just delivery mechanisms for pronunciation modeling but are also catalysts for learner autonomy and sustained oral engagement. The cross-cultural adaptability of AI tools across varied linguistic backgrounds and learning environments further strengthens their educational relevance. Whether in formal classrooms or informal self-directed settings, learners integrated AI practice into their routines, demonstrating increased motivation and ownership over their oral language development. Across all dimensions—acoustic precision, fluency control, learner autonomy, and motivational engagement—AI speech assistants consistently contributed to positive learner outcomes. These results confirm their emerging role as effective, scalable, and pedagogically sound resources for fostering English oral proficiency in diverse global contexts.

RECOMMENDATIONS

Based on the findings of this systematic review, it is recommended that language educators, curriculum designers, and learners integrate AI speech assistants as complementary tools for developing English pronunciation and fluency in both formal and informal learning environments. These systems should be incorporated into speaking-focused modules to provide learners with consistent opportunities for autonomous oral practice, particularly in contexts where access to native speaker interaction or individualized feedback is limited. Educators are encouraged to design structured speaking tasks around speech assistant interaction—such as question-answer simulations, pronunciation drills, and storytelling prompts—to promote sustained engagement and measurable linguistic gains. Institutions should also prioritize learner training on how to interpret implicit feedback from these tools, emphasizing strategies such as reformulation, articulation planning, and pitch modulation. For optimal results, learners should engage with these tools regularly over extended periods, ideally in combination with visual-acoustic feedback (e.g., waveform or pitch tracking) to support prosodic control. Furthermore, developers and language technology designers are advised to enhance AI systems with learner-sensitive features, such as accent adaptability, customizable feedback options, and detailed performance logs, to better serve a diverse global user base. Cross-cultural usability should be considered in future applications, ensuring that AI tools accommodate varying linguistic needs and learner contexts while supporting equitable access and engagement in second language oral skill development.

REFERENCES

[1]. Abdullah Al, M., Md Masud, K., Mohammad, M., & Hosne Ara, M. (2024). Behavioral Factors in Loan Default Prediction A Literature Review On Psychological And Socioeconomic Risk Indicators. *American Journal of Advanced Technology and Engineering Solutions*, *4*(01), 43-70. https://doi.org/10.63125/0jwtbn29

[2]. Abdur Razzak, C., Golam Qibria, L., & Md Arifur, R. (2024). Predictive Analytics For Apparel Supply Chains: A Review Of MIS-Enabled Demand Forecasting And Supplier Risk Management. *American Journal of Interdisciplinary Studies*, *5*(04), 01–23. https://doi.org/10.63125/80dwy222

[3]. Adar, C., & Md, N. (2023). Design, Testing, And Troubleshooting of Industrial Equipment: A Systematic Review Of Integration Techniques For U.S. Manufacturing Plants. *Review of Applied Science and Technology*, *2*(01), 53-84. https://doi.org/10.63125/893et038

[4]. Alastuey, M. C. B. (2011). Synchronous-Voice Computer-Mediated Communication: Effects on Pronunciation. *CALICO Journal*, *28*(1), 1-20. https://doi.org/10.11139/cj.28.1.1-20

[5]. Anika Jahan, M., Md Soyeb, R., & Tahmina Akter, R. (2025). Strategic Use Of Engagement Marketing in Digital Platforms: A Focused Analysis Of Roi And Consumer Psychology. *Journal of Sustainable Development and Policy*, *1*(01), 170-197. https://doi.org/10.63125/hm96p734

[6]. Audhkhasi, K., Rosenberg, A., Sethy, A., Ramabhadran, B., & Kingsbury, B. (2017). ICASSP - End-to-end ASR-free keyword search from speech. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, *NA*(NA), 4840-4844. https://doi.org/10.1109/icassp.2017.7953076

[7]. Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2022). 'Look, I can speak correctly': learning vocabulary and pronunciation through websites equipped with automatic speech recognition technology. *Computer Assisted Language Learning*, *37*(5-6), 1335-1363. https://doi.org/10.1080/09588221.2022.2080230

[8]. Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In (Vol. NA, pp. 13-34). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.07bes

[9]. Borges, A. F. T., Giraud, A.-L., Mansvelder, H. D., & Linkenkaer-Hansen, K. (2017). Scale-free amplitude modulation of neuronal oscillations tracks comprehension of accelerated speech. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *38*(3), 710-722. https://doi.org/10.1523/jneurosci.1515-17.2017

[10]. Burgess, J., & Spencer, S. (2000). Phonology and Pronunciation in Integrated Language Teaching and Teacher Education. *System*, *28*(2), 191-215. https://doi.org/10.1016/s0346-251x(00)00007-5

[11]. Burns, A. (2006). Integrating research and professional development on pronunciation teaching in a national adult ESL program. *TESL Reporter*, *39*(2), 34-41. https://doi.org/NA

[12]. Cao, Q., & Hao, H. (2021). Optimization of Intelligent English Pronunciation Training System Based on Android Platform. *Complexity*, *2021*(1), 1-11. https://doi.org/10.1155/2021/5537101

[13]. Cao, R., Pu, X., Du, X., Yang, W., Wang, J., Guo, H., Zhao, S., Yuan, Z., Zhang, C., Li, C., & Wang, Z. L. (2018). Screen-Printed Washable Electronic Textiles as Self-Powered Touch/Gesture Tribo-Sensors for Intelligent Human-Machine Interaction. *ACS nano*, *12*(6), 5190-5196. https://doi.org/10.1021/acsnano.8b02477

[14]. Chapelle, C. A. (2009). A Hidden Curriculum in Language Textbooks: Are Beginning Learners of French at U.S. Universities Taught about Canada? *The Modern Language Journal*, *93*(2), 139-152. https://doi.org/10.1111/j.1540-4781.2009.00852.x

[15]. Cieri, C., Miller, D., & Walker, K. (2004). LREC - The Fisher Corpus: a Resource for the Next Generations of Speech-to-Text.

[16]. Dai, Y., & Wu, Z. (2021). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: a mixed-methods study. *Computer Assisted Language Learning*, *36*(5-6), 861-884. https://doi.org/10.1080/09588221.2021.1952272

[17]. Dai, Y., & Wu, Z. (2022). Mobile-assisted peer feedback on EFL pronunciation: Outcome effects, interactional processes, and shaping factors. *System*, *111*(NA), 102953-102953. https://doi.org/10.1016/j.system.2022.102953

[18]. Daniels, P., & Iwago, K. (2017). The suitability of cloud-based speech recognition engines for language learning. *The JALT CALL Journal*, *13*(3), 211-221. https://doi.org/10.29140/jaltcall.v13n3.220

[19]. Derwing, T. M., & Munro, M. J. (2005). Second Language Accent and Pronunciation Teaching: A Research- Based Approach. *TESOL Quarterly*, *39*(3), 379-397. https://doi.org/10.2307/3588486

[20]. Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals : evidence-based perspectives for L2 teaching and research*. John Benjamins Publishing Company. https://doi.org/10.1075/lllt.42

[21]. Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does Popular Speech Recognition Software Work with ESL Speech. *TESOL Quarterly*, *34*(3), 592-603. https://doi.org/10.2307/3587748

[22]. Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J. (2014). Opening the Window on Comprehensible Pronunciation After 19 Years: A Workplace Training Study. *Language Learning*, *64*(3), 526-548. https://doi.org/10.1111/lang.12053

[23]. Derwing, T. M., & Rossiter, M. J. (2003). The Effects of Pronunciation Instruction on the Accuracy, Fluency, and Complexity of L2 Accented Speech. *Applied Language Learning*, *13*(1), 1-17. https://doi.org/NA
[24]. Derwing, T. M., Rossiter, M. J., & Munro, M. J. (2002). Teaching Native Speakers to Listen to Foreign-accented Speech. *Journal of Multilingual and Multicultural Development*, *23*(4), 245-259. https://doi.org/10.1080/01434630208666468
[25]. Dileep, A. D., & Sekhar, C. C. (2014). GMM-based intermediate matching kernel for classification of varying length patterns of long duration speech using support vector machines. *IEEE transactions on neural networks and learning systems*, *25*(8), 1421-1432. https://doi.org/10.1109/tnnls.2013.2293512
[26]. Dlaska, A., & Krekeler, C. (2008). Self-Assessment of Pronunciation. *System*, *36*(4), 506-516. https://doi.org/10.1016/j.system.2008.03.003
[27]. Dlaska, A., & Krekeler, C. (2013). The Short-Term Effects of Individual Corrective Feedback on L2 Pronunciation. *System*, *41*(1), 25-37. https://doi.org/10.1016/j.system.2013.01.005
[28]. Foote, J. A., Holtby, A. K., & Derwing, T. M. (2012). Survey of the Teaching of Pronunciation in Adult ESL Programs in Canada, 2010. *TESL Canada Journal*, *29*(1), 1-22. https://doi.org/10.18806/tesl.v29i1.1086
[29]. Golam Qibria, L., & Takbir Hossen, S. (2023). Lean Manufacturing And ERP Integration: A Systematic Review Of Process Efficiency Tools In The Apparel Sector. *American Journal of Scholarly Research and Innovation*, *2*(01), 104-129. https://doi.org/10.63125/mx7j4p06
[30]. Guion, S. G., & Pederson, E. (2007). Language Experience in Second Language Speech Learning: In honor of James Emil Flege - Investigating the role of attention in phonetic learning. In (Vol. NA, pp. 57-77). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.09gui
[31]. He, X., Zi, Y., Yu, H., Zhang, S. L., Wang, J., Ding, W., Zou, H., Zhang, W., Lu, C., & Wang, Z. L. (2017). An ultrathin paper-based self-powered system for portable electronics and wireless human-machine interaction. *Nano Energy*, *39*(NA), 328-336. https://doi.org/10.1016/j.nanoen.2017.06.046
[32]. Hosne Ara, M., Tonmoy, B., Mohammad, M., & Md Mostafizur, R. (2022). AI-ready data engineering pipelines: a review of medallion architecture and cloud-based integration models. *American Journal of Scholarly Research and Innovation*, *1*(01), 319-350. https://doi.org/10.63125/51kxtf08
[33]. Hu, X., Ackermann, H., Martin, J. A., Erb, M., Winkler, S., & Reiterer, S. M. (2012). Language aptitude for pronunciation in advanced second language (L2) Learners: Behavioural predictors and neural substrates. *Brain and language*, *127*(3), 366-376. https://doi.org/10.1016/j.bandl.2012.11.006
[34]. Hughes, H. G. A. (2004). The Cambridge Encyclopedia of the English Language (2nd edition). *Reference Reviews*, *18*(3), 28-29. https://doi.org/10.1108/09504120410528126
[35]. Istiaque, M., Dipon Das, R., Hasan, A., Samia, A., & Sayer Bin, S. (2023). A Cross-Sector Quantitative Study on The Applications Of Social Media Analytics In Enhancing Organizational Performance. *American Journal of Scholarly Research and Innovation*, *2*(02), 274-302. https://doi.org/10.63125/d8ree044
[36]. Istiaque, M., Dipon Das, R., Hasan, A., Samia, A., & Sayer Bin, S. (2024). Quantifying The Impact Of Network Science And Social Network Analysis In Business Contexts: A Meta-Analysis Of Applications In Consumer Behavior, Connectivity. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 58-89. https://doi.org/10.63125/vgkwe938
[37]. Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, *23*(1), 83-103. https://doi.org/10.1093/applin/23.1.83
[38]. Kang, O., & Rubin, D. L. (2009). Reverse Linguistic Stereotyping: Measuring the Effect of Listener Expectations on Speech Evaluation. *Journal of Language and Social Psychology*, *28*(4), 441-456. https://doi.org/10.1177/0261927x09341950
[39]. Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., Senft, E., & Beloaeme, T. (2017). HRI - Child Speech Recognition in Human-Robot Interaction: Evaluations and Recommendations. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, *NA*(NA), 82-90. https://doi.org/10.1145/2909824.3020229
[40]. Kennedy, S., & Trofimovich, P. (2010). Language awareness and second language pronunciation: a classroom study. *Language Awareness*, *19*(3), 171-185. https://doi.org/10.1080/09658416.2010.486439
[41]. Khan, A. S., Akter, M., Enni, M. A., & Khan, S. F. (2025). An in silico approach for the identification of detrimental missense SNPs and their potential impacts on human CRY2 protein. *Journal of Bangladesh Academy of Sciences*, *49*(1), 57-72. https://doi.org/10.3329/jbas.v49i1.71914
[42]. Kibishi, H., Hirabayashi, K., & Nakagawa, S. (2014). A statistical method of evaluating the pronunciation proficiency/intelligibility of English presentations by Japanese speakers. *ReCALL*, *27*(1), 58-83. https://doi.org/10.1017/s0958344014000251
[43]. Kutub Uddin, A., Md Mostafizur, R., Afrin Binta, H., & Maniruzzaman, B. (2022). Forecasting Future Investment Value with Machine Learning, Neural Networks, And Ensemble Learning: A Meta-Analytic Study. *Review of Applied Science and Technology*, *1*(02), 01-25. https://doi.org/10.63125/edxgjg56
[44]. Lasi, F. (2020). A Study on the Ability of Supra-Segmental and Segmental Aspects in English Pronunciation. *Ethical Lingua: Journal of Language Teaching and Literature*, *7*(2), 426-437. https://doi.org/10.30605/25409190.222

[45]. LeVelle, K., & Levis, J. M. (2014). Social Dynamics in Second Language Accent - 5. Understanding the Impact of Social Factors on L2 Pronunciation: Insights from Learners. In (Vol. NA, pp. 97-118). DE GRUYTER. https://doi.org/10.1515/9781614511762.97
[46]. MacKay, I. R. A., & Flege, J. E. (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. *Applied Psycholinguistics*, *25*(3), 373-396. https://doi.org/10.1017/s0142716404001171
[47]. Mansura Akter, E. (2023). Applications Of Allele-Specific PCR In Early Detection of Hereditary Disorders: A Systematic Review Of Techniques And Outcomes. *Review of Applied Science and Technology*, *2*(03), 1-26. https://doi.org/10.63125/n4h7t156
[48]. Mansura Akter, E. (2025). Bioinformatics-Driven Approaches in Public Health Genomics: A Review Of Computational SNP And Mutation Analysis. *International Journal of Scientific Interdisciplinary Research*, *6*(1), 88-118. https://doi.org/10.63125/e6pxkn12
[49]. Mansura Akter, E., & Md Abdul Ahad, M. (2022). In Silico drug repurposing for inflammatory diseases: a systematic review of molecular docking and virtual screening studies. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 35-64. https://doi.org/10.63125/j1hbts51
[50]. Mansura Akter, E., & Shaiful, M. (2024). A systematic review of SNP polymorphism studies in South Asian populations: implications for diabetes and autoimmune disorders. *American Journal of Scholarly Research and Innovation*, *3*(01), 20-51. https://doi.org/10.63125/8nvxcb96
[51]. Martin, S., Alvarez, I. M., & Espasa, A. (2022). Video feedback and Foreign Language Anxiety in online pronunciation tasks. *International Journal of Educational Technology in Higher Education*, *19*(1), NA-NA. https://doi.org/10.1186/s41239-022-00324-y
[52]. McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of Automatic Speech Recognition. *System*, *57*(NA), 25-42. https://doi.org/10.1016/j.system.2015.12.013
[53]. McCrocklin, S. (2019). Learners' Feedback Regarding ASR-based Dictation Practice for Pronunciation Learning. *CALICO Journal*, *36*(2), 119-137. https://doi.org/10.1558/cj.34738
[54]. Md Arafat, S., Md Imran, K., Hasib, A., Md Jobayer Ibne, S., & Md Sanjid, K. (2025). Investigating Key Attributes for Circular Economy Implementation In Manufacturing Supply Chains: Impacts On The Triple Bottom Line. *Review of Applied Science and Technology*, *4*(02), 145-175. https://doi.org/10.63125/fnsy0e41
[55]. Md Atiqur Rahman, K., Md Abdur, R., Niger, S., & Mst Shamima, A. (2025). Development Of a Fog Computing-Based Real-Time Flood Prediction And Early Warning System Using Machine Learning And Remote Sensing Data. *Journal of Sustainable Development and Policy*, *1*(01), 144-169. https://doi.org/10.63125/6y0qwr92
[56]. Md Jakaria, T., Md, A., Zayadul, H., & Emdadul, H. (2025). Advances In High-Efficiency Solar Photovoltaic Materials: A Comprehensive Review of Perovskite And Tandem Cell Technologies. *American Journal of Advanced Technology and Engineering Solutions*, *1*(01), 201-225. https://doi.org/10.63125/5amnvb37
[57]. Md Mahamudur Rahaman, S. (2022). Electrical And Mechanical Troubleshooting in Medical And Diagnostic Device Manufacturing: A Systematic Review Of Industry Safety And Performance Protocols. *American Journal of Scholarly Research and Innovation*, *1*(01), 295-318. https://doi.org/10.63125/d68y3590
[58]. Md Masud, K., Sazzad, I., Mohammad, M., & Noor Alam, S. (2025). Digitization In Retail Banking: A Review of Customer Engagement And Financial Product Adoption In South Asia. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 42-46. https://doi.org/10.63125/cv50rf30
[59]. Md, N., Golam Qibria, L., Abdur Razzak, C., & Khan, M. A. M. (2025). Predictive Maintenance In Power Transformers: A Systematic Review Of AI And IOT Applications. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 34-47. https://doi.org/10.63125/r72yd809
[60]. Md Nazrul Islam, K., & Debashish, G. (2025). Cybercrime and contractual liability: a systematic review of legal precedents and risk mitigation frameworks. *Journal of Sustainable Development and Policy*, *1*(01), 01-24. https://doi.org/10.63125/x3cd4413
[61]. Md Nazrul Islam, K., & Ishtiaque, A. (2025). A systematic review of judicial reforms and legal access strategies in the age of cybercrime and digital evidence. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 01-29. https://doi.org/10.63125/96ex9767
[62]. Md Nur Hasan, M., Md Musfiqur, R., & Debashish, G. (2022). Strategic Decision-Making in Digital Retail Supply Chains: Harnessing AI-Driven Business Intelligence From Customer Data. *Review of Applied Science and Technology*, *1*(03), 01-31. https://doi.org/10.63125/6a7rpy62
[63]. Md Takbir Hossen, S., Abdullah Al, M., Siful, I., & Md Mostafizur, R. (2025). Transformative applications of ai in emerging technology sectors: a comprehensive meta-analytical review of use cases in healthcare, retail, and cybersecurity. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 121-141. https://doi.org/10.63125/45zpb481
[64]. Md Takbir Hossen, S., Ishtiaque, A., & Md Atiqur, R. (2023). AI-Based Smart Textile Wearables For Remote Health Surveillance And Critical Emergency Alerts: A Systematic Literature Review. *American Journal of Scholarly Research and Innovation*, *2*(02), 1-29. https://doi.org/10.63125/ceqapd08

[65]. Md Takbir Hossen, S., & Md Atiqur, R. (2022). Advancements In 3d Printing Techniques For Polymer Fiber-Reinforced Textile Composites: A Systematic Literature Review. *American Journal of Interdisciplinary Studies*, *3*(04), 32-60. https://doi.org/10.63125/s4r5m391
[66]. Md Tawfiqul, I., Meherun, N., Mahin, K., & Mahmudur Rahman, M. (2022). Systematic Review of Cybersecurity Threats In IOT Devices Focusing On Risk Vectors Vulnerabilities And Mitigation Strategies. *American Journal of Scholarly Research and Innovation*, *1*(01), 108-136. https://doi.org/10.63125/wh17mf19
[67]. Moyer, A. (2014). Social Dynamics in Second Language Accent - 1. The Social Nature of L2 Pronunciation. In (Vol. NA, pp. 11-30). DE GRUYTER. https://doi.org/10.1515/9781614511762.11
[68]. Mst Shamima, A., Niger, S., Md Atiqur Rahman, K., & Mohammad, M. (2023). Business Intelligence-Driven Healthcare: Integrating Big Data And Machine Learning For Strategic Cost Reduction And Quality Care Delivery. *American Journal of Interdisciplinary Studies*, *4*(02), 01-28. https://doi.org/10.63125/crv1xp27
[69]. Murphy, J. M., & Baker, A. L. (2015). *The Handbook of English Pronunciation - History of ESL Pronunciation Teaching* (Vol. NA). Wiley. https://doi.org/10.1002/9781118346952.ch3
[70]. Pinget, A.-F., Bosker, H. R., Quené, H., & de Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, *31*(3), 349-365. https://doi.org/10.1177/0265532214526177
[71]. Potamianos, A., & Narayanan, S. S. (2003). Robust recognition of children's speech. *IEEE Transactions on Speech and Audio Processing*, *11*(6), 603-616. https://doi.org/10.1109/tsa.2003.818026
[72]. Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N. K., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., & Vesely, K. (2011). The Kaldi Speech Recognition Toolkit.
[73]. Qian, Y., Ubale, R., Lange, P., Evanini, K., & Soong, F. K. (2018). ISCSLP - From Speech Signals to Semantics — Tagging Performance at Acoustic, Phonetic and Word Levels. *2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, *NA*(NA), 280-284. https://doi.org/10.1109/iscslp.2018.8706581
[74]. Rezwanul Ashraf, R., & Hosne Ara, M. (2023). Visual communication in industrial safety systems: a review of UI/UX design for risk alerts and warnings. *American Journal of Scholarly Research and Innovation*, *2*(02), 217-245. https://doi.org/10.63125/wbv4z521
[75]. Sanjai, V., Sanath Kumar, C., Maniruzzaman, B., & Farhana Zaman, R. (2023). Integrating Artificial Intelligence in Strategic Business Decision-Making: A Systematic Review Of Predictive Models. *International Journal of Scientific Interdisciplinary Research*, *4*(1), 01-26. https://doi.org/10.63125/s5skge53
[76]. Sanjai, V., Sanath Kumar, C., Sadia, Z., & Rony, S. (2025). Ai And Quantum Computing For Carbon-Neutral Supply Chains: A Systematic Review Of Innovations. *American Journal of Interdisciplinary Studies*, *6*(1), 40-75. https://doi.org/10.63125/nrdx7d32
[77]. Sardegna, V. G., Lee, J., & Kusey, C. (2017). Self-Efficacy, Attitudes, and Choice of Strategies for English Pronunciation Learning. *Language Learning*, *68*(1), 83-114. https://doi.org/10.1111/lang.12263
[78]. Sazzad, I. (2025a). Public Finance and Policy Effectiveness A Review Of Participatory Budgeting In Local Governance Systems. *Journal of Sustainable Development and Policy*, *1*(01), 115-143. https://doi.org/10.63125/p3p09p46
[79]. Sazzad, I. (2025b). A Systematic Review of Public Budgeting Strategies In Developing Economies: Tools For Transparent Fiscal Governance. *American Journal of Advanced Technology and Engineering Solutions*, *1*(01), 602-635. https://doi.org/10.63125/wm547117
[80]. Sazzad, I., & Md Nazrul Islam, K. (2022). Project impact assessment frameworks in nonprofit development: a review of case studies from south asia. *American Journal of Scholarly Research and Innovation*, *1*(01), 270-294. https://doi.org/10.63125/eeja0t77
[81]. Shaiful, M., & Mansura Akter, E. (2025). AS-PCR In Molecular Diagnostics: A Systematic Review of Applications In Genetic Disease Screening. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 98-120. https://doi.org/10.63125/570jb007
[82]. Shen, D., & Zhao, W. (2023). A Method for Improving the Pronunciation Quality of Vocal Music Students Based on Big Data Technology. *International Journal of Web-Based Learning and Teaching Technologies*, *19*(1), 1-18. https://doi.org/10.4018/ijwltt.335034
[83]. Shi, Q., & Lee, C. (2019). Self-Powered Bio-Inspired Spider-Net-Coding Interface Using Single-Electrode Triboelectric Nanogenerator. *Advanced science (Weinheim, Baden-Wurttemberg, Germany)*, *6*(15), 1900617-1900617. https://doi.org/10.1002/advs.201900617
[84]. Smit, U., & Dalton, C. (2000). Motivational patterns in advanced EFL pronunciation learners. *IRAL - International Review of Applied Linguistics in Language Teaching*, *38*(3-4), 229-246. https://doi.org/10.1515/iral.2000.38.3-4.229
[85]. Sohel, R., & Md, A. (2022). A Comprehensive Systematic Literature Review on Perovskite Solar Cells: Advancements, Efficiency Optimization, And Commercialization Potential For Next-Generation Photovoltaics. *American Journal of Scholarly Research and Innovation*, *1*(01), 137-185. https://doi.org/10.63125/843z2648

[86]. Subrato, S. (2018). Resident's Awareness Towards Sustainable Tourism for Ecotourism Destination in Sundarban Forest, Bangladesh. *Pacific International Journal*, *1*(1), 32-45. https://doi.org/10.55014/pij.v1i1.38
[87]. Subrato, S. (2025). Role of management information systems in environmental risk assessment: a systematic review of geographic and ecological applications. *American Journal of Interdisciplinary Studies*, 6(1), 95–126. https://doi.org/10.63125/k27tnn83
[88]. Subrato, S., & Faria, J. (2025). AI-driven MIS applications in environmental risk monitoring: a systematic review of predictive geographic information systems. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 81-97. https://doi.org/10.63125/pnx77873
[89]. Tahmina Akter, R. (2025). AI-driven marketing analytics for retail strategy: a systematic review of data-backed campaign optimization. *International Journal of Scientific Interdisciplinary Research*, 6(1), 28-59. https://doi.org/10.63125/0k4k5585
[90]. Tahmina Akter, R., & Abdur Razzak, C. (2022). The Role Of Artificial Intelligence In Vendor Performance Evaluation Within Digital Retail Supply Chains: A Review Of Strategic Decision-Making Models. *American Journal of Scholarly Research and Innovation*, *1*(01), 220-248. https://doi.org/10.63125/96jj3j86
[91]. Tahmina Akter, R., Debashish, G., Md Soyeb, R., & Abdullah Al, M. (2023). A Systematic Review of AI-Enhanced Decision Support Tools in Information Systems: Strategic Applications In Service-Oriented Enterprises And Enterprise Planning. *Review of Applied Science and Technology*, *2*(01), 26-52. https://doi.org/10.63125/73djw422
[92]. Tahmina Akter, R., Md Arifur, R., & Anika Jahan, M. (2024). Customer relationship management and data-driven decision-making in modern enterprises: a systematic literature review. *American Journal of Advanced Technology and Engineering Solutions*, *4*(04), 57-82. https://doi.org/10.63125/jetvam38
[93]. Tan, H., Tao, Q., Pande, I., Majumdar, S., Liu, F., Zhou, Y., Persson, P., Rosen, J., & van Dijken, S. (2020). Tactile sensory coding and learning with bio-inspired optoelectronic spiking afferent nerves. *Nature communications*, *11*(1), 1369-1369. https://doi.org/10.1038/s41467-020-15105-2
[94]. Tetariy, E., Gishri, M., Har-Lev, B., Aharonson, V., & Moyal, A. (2012). An efficient lattice-based phonetic search method for accelerating keyword spotting in large speech databases. *International Journal of Speech Technology*, *16*(2), 161-169. https://doi.org/10.1007/s10772-012-9171-3
[95]. Trofimovich, P., & Baker, W. (2006). Learning Second Language Suprasegmentals: Effect of L2 Experience on Prosody and Fluency Characteristics of L2 Speech. *Studies in Second Language Acquisition*, *28*(1), 1-30. https://doi.org/10.1017/s0272263106060013
[96]. van Doremalen, J., Boves, L., Colpaert, J., Cucchiarini, C., & Strik, H. (2016). Evaluating Automatic Speech Recognition-Based Language Learning Systems: A Case Study. *Computer Assisted Language Learning*, *29*(4), 833-851. https://doi.org/10.1080/09588221.2016.1167090
[97]. van Loon, J. (2002). Improving Pronunciation of Adult ESL Students. *TESL Canada Journal*, *20*(1), 83-88. https://doi.org/10.18806/tesl.v20i1.941
[98]. Wang, K., Lou, Z., Wang, L., Zhao, L., Zhao, S., Wang, D., Han, W., Jiang, K., & Shen, G. (2019). Bioinspired Interlocked Structure-Induced High Deformability for Two-Dimensional Titanium Carbide (MXene)/Natural Microcapsule-Based Flexible Pressure Sensors. *ACS nano*, *13*(8), 9139-9147. https://doi.org/10.1021/acsnano.9b03454
[99]. Wang, X., Zhang, Y., Zhang, X., Huo, Z., Li, X., Que, M., Peng, Z., Wang, H., & Pan, C. (2018). A Highly Stretchable Transparent Self-Powered Triboelectric Tactile Sensor with Metallized Nanofibers for Wearable Electronics. *Advanced materials (Deerfield Beach, Fla.)*, *30*(12), 1706738-NA. https://doi.org/10.1002/adma.201706738
[100]. Wen, F., Sun, Z., He, T., Shi, Q., Zhu, M., Zhang, Z., Li, L., Zhang, T., & Lee, C. (2020). Machine Learning Glove Using Self-Powered Conductive Superhydrophobic Triboelectric Textile for Gesture Recognition in VR/AR Applications. *Advanced science (Weinheim, Baden-Wurttemberg, Germany)*, *7*(14), 2000261-2000261. https://doi.org/10.1002/advs.202000261
[101]. Werker, J. F., & Tees, R. C. (2002). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *25*(1), 121-133. https://doi.org/10.1016/s0163-6383(02)00093-0
[102]. Xiong, Y., Shen, Y., Tian, L., Hu, Y., Zhu, P., Sun, R., & Wong, C.-P. (2020). A flexible, ultra-highly sensitive and stable capacitive pressure sensor with convex microarrays for motion and health monitoring. *Nano Energy*, *70*(NA), 104436-NA. https://doi.org/10.1016/j.nanoen.2019.104436
[103]. Yao, S., Ren, P., Song, R., Liu, Y., Huang, Q., Dong, J., O'Connor, B., & Zhu, Y. (2019). Nanomaterial-Enabled Flexible and Stretchable Sensing Systems: Processing, Integration, and Applications. *Advanced materials (Deerfield Beach, Fla.)*, *32*(15), 1902343-NA. https://doi.org/10.1002/adma.201902343
[104]. Zhao, J., Bu, T., Zhang, X., Pang, Y., Li, W., Zhang, Z., Liu, G., Wang, Z. L., & Zhang, C. (2020). Intrinsically Stretchable Organic-Tribotronic-Transistor for Tactile Sensing. *Research (Washington, D.C.)*, *2020*(NA), 1398903-NA. https://doi.org/10.34133/2020/1398903

[105]. Zhu, J., Zhang, X., & Li, J. (2022). Using AR filters in L2 pronunciation training: practice, perfection, and willingness to share. *Computer Assisted Language Learning*, *37*(5-6), 1364-1396. https://doi.org/10.1080/09588221.2022.2080716

[106]. Zhu, M., Sun, Z., Zhang, Z., Shi, Q., He, T., Liu, H., Chen, T., & Lee, C. (2020). Haptic-feedback smart glove as a creative human-machine interface (HMI) for virtual/augmented reality applications. *Science advances*, 6(19), eaaz8693-NA. https://doi.org/10.1126/sciadv.aaz8693