

Article

Developing an AI-Powered Pronunciation Application to Improve English Pronunciation of Thai ESP Learners

Jiraporn Lao-un and Dararat Khampusaen * 

Faculty of Humanities and Social Sciences, Khon Kaen University, Khon Kaen 40002, Thailand;
jirapornl@kkumail.com

* Correspondence: darkha@kku.ac.th

Abstract

This study examined the effects of using specially designed AI-mediated pronunciation application in enhancing the production of English fricative consonants among Thai English for Specific Purposes (ESP) learners. The research utilized a quasi-experimental design involving intact classes of 74 undergraduate students majoring in Thai Dance and Music Education, divided into control ($N = 38$) and experimental ($N = 36$) groups. Grounded in Skill Acquisition Theory, the experimental group received pronunciation training via a custom-designed AI application leveraging automatic speech recognition (ASR), offering ESP contextualized practices, real-time, and individualized feedback. In contrast, the control group underwent traditional teacher-led articulatory and teacher-assisted feedback. Pre- and post-test evaluations measured pronunciation for nine target fricatives in ESP-relevant contexts. The statistical analyses revealed significant improvements in both groups, with the AI-mediated group demonstrating substantially greater gains, particularly on challenging sounds absent in Thai, such as /θ/, /ð/, /z/, /ʃ/, and /h/. The findings underscore the potential of AI-driven interventions to address language-specific phonological challenges through personalized, immediate feedback and adaptive practices. The study provides empirical evidence for integrating advanced technology into ESP pronunciation pedagogy, informing future curriculum design for EFL contexts. Implications for theory, practice, and future research are discussed, emphasizing tailored technological solutions for language learners with specific phonological profiles.

Keywords: AI-mediated pronunciation; English for specific purposes; Thai EFL learners; skill acquisition theory; automatic speech recognition



Academic Editor: Jette G. Hansen Edwards

Received: 9 June 2025

Revised: 6 October 2025

Accepted: 14 October 2025

Published: 28 October 2025

Citation: Lao-un, J., & Khampusaen, D. (2025). Developing an AI-Powered Pronunciation Application to Improve English Pronunciation of Thai ESP Learners. *Languages*, 10(11), 273. <https://doi.org/10.3390/languages10110273>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

English proficiency has become a crucial skill in today's globalized world, with pronunciation playing a pivotal role in effective communication. For non-native English speakers, particularly those in English as a Foreign Language (EFL) contexts, mastering English pronunciation presents significant challenges due to limited exposure to authentic language input and interference from first-language phonological systems (Derwing & Munro, 2015; Levis, 2018). In foreign contexts where English is the lingua franca, these pronunciation issues frequently endure despite years of formal education, impeding successful communication (Jenkins, 2000; Seidlhofer, 2011).

In Thailand, where English is taught as a foreign language, pronunciation issues are especially noticeable due to the significant phonological variations between the Thai and English sound systems. Thai learners frequently struggle with several English phonemes

absent in their native language. These difficulties are often compounded by the limited opportunities for authentic language exposure in the Thai EFL environment, where learners typically encounter English only within classroom settings (Khamkhien, 2010; Noom-ura, 2013). Moreover, English teachers in Thailand often lack specialized training in pronunciation pedagogy, leading to inadequate or inconsistent pronunciation instruction (Cedar & Termjai, 2021; Wongsuriya, 2020).

Traditional English pronunciation pedagogy has shown limited effectiveness in addressing these challenges as the focus often relies on repetition drills and teacher modeling that do not provide the personalized, immediate feedback necessary for sustained improvement (Thomson & Derwing, 2015). Furthermore, currently, pronunciation teaching materials in Thai ESP contexts are frequently outdated, inadequately addressing occupation-specific communication needs and lacking authentic input pertinent to learners' professional domains (Sukying et al., 2023). These challenges are exacerbated by large class sizes in Thai educational institutions, which severely limit opportunities for individualized pronunciation feedback (Chanaroke & Niemprapan, 2020; Khamkhien, 2010; Noom-ura, 2013). As is generally known, Thai teachers often prioritize grammar and vocabulary over pronunciation, partly due to time constraints and partly because pronunciation is not heavily weighted in standardized English examinations (Chanaroke & Niemprapan, 2020). Consequently, many Thai ESP graduates enter the workforce with pronunciation difficulties that impair their communicative effectiveness in professional contexts where English is required (Khampusaen et al., 2023).

Recent technological advances, particularly in artificial intelligence (AI), offer promising alternatives to promising supplements to teacher-led instruction. It is mentioned that AI-mediated pronunciation training systems can provide individualized feedback, detect specific pronunciation errors, and offer targeted remediation tailored to a learner's first language background and specific difficulties (Guskaroska, 2019; T. Schmidt & Strassner, 2022). As a result, AI has been considered a potential tool to promote the pronunciation skills of EFL learners in the Thai context. Furthermore, a growing body of research examines the effectiveness of computer-assisted pronunciation training (CAPT) in various contexts (Bashori et al., 2024; Lara et al., 2024; McCrocklin et al., 2022; Mohammadkarimi, 2024; Rajkumari et al., 2024). the studies focusing specifically on AI-based applications for Thai ESP learners remain limited. These unique challenges pose a barrier to integrating AI to assist in language learning. Due to its distinctive phonological system and the specific professional needs of ESP learners. This research gap necessitates empirical investigation to determine the potential benefits of integrating AI technology into pronunciation instruction for this specific population.

The present study addresses this gap by examining the extent to which AI-mediated pronunciation training affects Thai ESP learners' pronunciation of English fricative consonants. By comparing the effectiveness of AI-based instruction with traditional methods, this research aims to contribute to our understanding of how technological innovations can enhance pronunciation pedagogy for learners from specific linguistic backgrounds. The findings hold important implications for language teaching practices, curriculum design, and the development of technology-enhanced learning materials for ESP contexts in Thailand and beyond.

2. Literature Review

2.1. Pronunciation Challenges for Thai Learners of English

The literature review examines the growing body of research on artificial intelligence (AI) tools for improving English pronunciation skills. The Thai and English phonological systems differ significantly, creating specific challenges for Thai learners of English

pronunciation. Thai contains fewer consonant phonemes (21) compared to English (24), and critically lacks several English fricatives (Kanokpermpoon, 2007; Naruemon et al., 2024). The absence of interdental fricatives (/θ/ and /ð/), voiced fricatives (/v/, /z/, and /ʒ/), and certain distinctions between similar sounds in the Thai phonological inventory frequently results in substitution errors, where Thai learners replace unfamiliar English phonemes with the closest available Thai sound. English fricatives /v/, /θ/, /ð/, /z/, /ʃ/, and /ʒ/ do not exist in Thai. This leads to Thai learners often substituting /v/ with /f/, /θ/ and /ð/ with /t/ or /s/, /z/ with /s/, /ʃ/ with /s/ or /tʃ/, as well as /dʒ/ with /tɕ/ (Kanokpermpoon, 2007; Kitikanan, 2017; Naruemon et al., 2024).

There are a number of research studies addressing specific pronunciation difficulties in the Thai context (Behr, 2022; Huang & Kitikanan, 2022; Kitikanan, 2017; Naruemon et al., 2024). For example, research by Naruemon et al. (2024) demonstrated that Thai university students particularly struggle with English fricatives, often replacing /θ/ with /t/ or /s/, /ð/ with /d/, /v/ with /w/ or /f/, /z/ with /s/, and /ʒ/ with /ch/ or /s/. These substitutions can lead to communication breakdowns, especially in ESP contexts where precision in pronunciation may be crucial for professional effectiveness (Jenkins, 2000; Peerachachayane, 2022). Furthermore, Thai is a tonal language with significantly different stress patterns from English. The syllable-timed rhythm of Thai contrasts with the stress-timed pattern of English, creating additional pronunciation difficulties (Naruemon et al., 2024).

2.2. L2 Pronunciation Learning Theory

2.2.1. Speech Learning Model (SLM)

The Speech Learning Model (SLM) provides a comprehensive framework for understanding cross-linguistic segmental acquisition patterns, establishing the theoretical foundation for examining Thai learners' acquisition of English fricatives. This model emphasizes the fundamental role of native language (L1) phonetic categories in shaping second language (L2) speech perception and production, suggesting that L2 learners often assimilate new sounds to their existing L1 categories, which consequently leads to difficulties in accurate pronunciation and comprehension (Flege, 1995).

Building upon the original framework, the revised Speech Learning Model (SLM-r) articulates several key principles that directly inform the current investigation of Thai ESP learners. The model fundamentally establishes that L2 learners' pronunciation accuracy is intrinsically linked to their perceptual abilities. According to Flege and Bohn (2021), learners who cannot reliably distinguish between L2 sounds and their closest L1 counterparts will experience persistent production difficulties. This principle has direct implications for Thai learners acquiring English fricatives, given that many English fricative sounds either do not exist in Thai or occupy different phonetic spaces within the sound system.

The Perceived Dissimilarity Principle, which forms the core of SLM-r's theoretical framework, offers critical predictions regarding phonetic category formation. This principle predicts that the greater the perceived phonetic dissimilarity between an English fricative and the closest Thai sound, the more likely Thai learners are to successfully establish a new phonetic category. Research by Kitikanan (2017) demonstrates how this principle explains the differential success Thai learners experience with various English fricatives. For instance, interdental fricatives /θ/ and /ð/, which have no Thai equivalents, may actually be easier to acquire than sounds like /v/, which Thai learners might assimilate to existing L1 categories such as /w/ or /f/.

The Position-Sensitive Allophone Mapping principle within SLM-r provides additional insights into context-dependent sound production patterns. This principle suggests that learners map L2 sounds to L1 sounds at the level of context-sensitive variants rather than

abstract phonemes. This concept holds particular relevance for Thai ESP learners, as English fricatives may be realized differently across phonetic contexts, including initial, medial, and final positions. Thai phonotactic constraints may further influence how learners perceive and produce these sounds in various word positions (Isarankura, 2016; Kanokpermpoon, 2007; Naruemon et al., 2024). The theoretical predictions of SLM-r align closely with documented pronunciation challenges faced by Thai learners in previous research. Previous studies by Naruemon et al. (2024) and Kanokpermpoon (2007) demonstrate that Thai speakers frequently substitute /θ/ with /t/ or /s/, /ð/ with /d/, /v/ with /w/ or /f/, and /z/ with /s/. These substitution patterns can be understood through SLM-r's framework as instances where perceived similarity between L1 and L2 sounds prevents the formation of distinct L2 phonetic categories.

The model's emphasis on L1-L2 phonetic interactions provides theoretical grounding for understanding the effectiveness of AI-mediated pronunciation training. When learners successfully form new phonetic categories for English fricatives, these categories may dissimilate from neighboring Thai categories to maintain phonetic contrast. Conversely, unsuccessful category formation leads to a merger between Thai and English phonetic categories, resulting in persistent substitution errors. The immediate and consistent feedback provided by AI applications may facilitate the formation of distinct L2 categories by enhancing learners' awareness of phonetic differences that might otherwise remain unnoticed. This enhanced awareness represents a crucial mechanism through which technology-mediated instruction can address the perceptual foundations of pronunciation difficulties.

2.2.2. Skill Acquisition Theory (SAT)

The present study aligns with Skill Acquisition Theory (SAT), a cognitive framework that offers an explanation for how language skills, including pronunciation, develop through stages toward automaticity. Originally developed by Anderson (1982) as the Adaptive Control of Thought (ACT) model and later applied to second language acquisition by R. DeKeyser (2007b), SAT provides a comprehensive account of how learners progress from conscious knowledge of language forms to automatic performance (Anderson, 1992; R. DeKeyser, 2007b). According to SAT, skill development progresses through three primary stages: cognitive, associative, and autonomous (R. M. DeKeyser, 1997; Taie, 2014). In the cognitive stage, learners develop declarative knowledge about pronunciation features, such as explicit understanding of how specific sounds are articulated. Thai learners of English, for instance, must first develop declarative knowledge about the articulatory features of English fricatives that do not exist in their L1 phonological system, such as the interdental fricatives /θ/ and /ð/ (Kormos, 2006).

During the associative stage, learners begin to proceduralize this declarative knowledge through consistent practice, gradually reducing error rates and processing time. For pronunciation skills, this involves repeated practice producing target sounds while receiving feedback on accuracy. The quality and immediacy of feedback during this stage are crucial for successful proceduralization (R. M. DeKeyser, 2015). At this stage, Thai learners might begin producing English fricatives with increasing accuracy but still require conscious attention to their articulation. Finally, in the autonomous stage, pronunciation skills become largely automatic, requiring minimal attentional resources. Learners can produce target sounds accurately in spontaneous speech without conscious monitoring (Segalowitz, 2010).

Skill Acquisition Theory (SAT) emphasizes that the transition from declarative to procedural knowledge requires extensive, deliberate practice with appropriate feedback (R. DeKeyser, 2007a). However, this emphasis on explicit declarative knowledge as a prerequisite for procedural development requires qualification in the context of pronuncia-

tion learning. Empirical evidence suggests that learners may develop implicit articulatory representations through exposure-based learning that operates independently of explicit instruction (Ellis, 2005; Rebuschat, 2013). This theoretical perspective is particularly relevant to AI-mediated pronunciation training, as such technology can provide the consistent, individualized feedback crucial for proceduralization. Additionally, usage-based approaches to L2 acquisition demonstrate that repeated exposure to target sounds in meaningful contexts can lead to procedural knowledge without extensive declarative instruction (Tomasello, 2003; Wulff & Ellis, 2018). Unlike traditional classroom settings where feedback may be inconsistent or delayed, AI applications can offer immediate, targeted feedback on specific articulation errors, potentially accelerating the proceduralization process while simultaneously supporting implicit learning through repeated exposure to target pronunciations (R. Schmidt, 2001; R. M. DeKeyser, 2015). Furthermore, SAT suggests that skill development is highly specific, with limited transfer between skills (R. M. DeKeyser, 2015). This specificity principle implies that practice should closely resemble the target performance conditions. For ESP learners, this means practicing pronunciation in contexts that mirror authentic professional communication scenarios. AI-mediated training can be designed to incorporate discipline-specific vocabulary and discourse patterns, aligning practice conditions with target performance requirements (Khampusaen et al., 2023).

The exponential learning curve proposed by SAT also has important implications for pronunciation instruction. According to this principle, learning gains are most rapid during initial practice stages but gradually diminish with continued practice (R. DeKeyser, 2007b). This suggests that AI-mediated pronunciation training might show the most significant effects during earlier learning phases, with progress becoming more incremental over time. The intervention period in the current study allows for observation of this learning trajectory. Moreover, SAT acknowledges the role of individual differences in skill acquisition. Factors such as working memory capacity, phonological awareness, and aptitude can influence how efficiently learners progress through the stages of skill development (Kormos, 2006; Skehan, 2015). AI-mediated pronunciation training may address these individual differences by adapting to learners' specific needs and learning rates, offering personalized practice opportunities not possible in traditional group instruction.

For Thai ESP learners acquiring English fricative pronunciation, SAT predicts that the systematic practice and immediate feedback provided by AI applications should facilitate the transition from declarative knowledge about fricative articulation to procedural knowledge, eventually leading to more automatic production (Derwing & Munro, 2015). The present study examines this prediction by comparing the effectiveness of AI-mediated training, which emphasizes consistent feedback and targeted practice, with traditional instruction using teacher-led methods, peer interaction, and teacher-assisted feedback.

2.3. AI-Mediated Applications in Pronunciation Teaching

With significant advancements occurring around 2005, artificial intelligence has introduced new possibilities for pronunciation instruction through machine learning algorithms that can analyze speech patterns, identify errors, and provide personalized feedback at a level of detail and consistency difficult to achieve in traditional classroom settings. Building on this foundation, technological tools such as speech recognition and accent analysis have become integral to improving EFL learners' pronunciation. These tools leverage advanced algorithms to analyze learners' speech patterns, identify errors, and provide corrective feedback. For instance, AI-assisted pronunciation correction tools use automatic speech recognition (ASR) to transcribe spoken input and compare it with target pronunciations, offering detailed feedback on phonemes, intonation, and stress points (Dennis, 2024; Khampusaen et al., 2023; Lara et al., 2024). Similarly, systems like ELSA Speak and NOVO employ

speech recognition to detect pronunciation errors and provide lessons tailored to individual needs (Aryanti & Santosa, 2024; Bashori et al., 2024)

Accordingly, research findings demonstrate that ASR technology significantly enhances EFL learners' pronunciation skills through immediate feedback mechanisms that help learners identify and correct pronunciation errors (Akhila et al., 2024; Dennis, 2024; McCrocklin et al., 2022). Research findings demonstrate that ASR technology can enhance EFL learners' pronunciation skills through immediate feedback mechanisms, though with important limitations. Current ASR systems may struggle with non-native accented speech and can provide inaccurate feedback on pronunciation that is perfectly comprehensible to human listeners (Ngo et al., 2024; Liu et al., 2025). Despite these limitations, tools like ELSA Speak and Listnr have shown improvements in segmental accuracy and overall communicative competence when used appropriately (Mohammadkarimi, 2024; McCrocklin et al., 2022).

One of the key advantages of ASR technology is its ability to provide personalized feedback. AI-assisted tools like SpeakBuddy and SalukiSpeech offer tailored exercises based on individual learner needs, adapting to their strengths and weaknesses over time (Akhila et al., 2024; McCrocklin et al., 2022). For example, SalukiSpeech uses machine learning models to evolve with user interactions, ensuring that feedback becomes more precise as learners progress (McCrocklin et al., 2022). This personalized approach has been shown to outperform traditional methods in improving pronunciation accuracy and fluency (Shafiee Rad & Roohani, 2024). Moreover, ASR-based systems excel in providing real-time evaluation of pronunciation performance. This real-time feedback is particularly effective in developing accurate pronunciation habits, as learners can adjust their speech on the spot (Kang et al., 2024). Additionally, some systems, such as the one developed by Lee et al. (2013), combine ASR with phonetic analysis to identify specific phonemes and intonation patterns that learners struggle with. Beyond technical capabilities, ASR technology enhances learner motivation and engagement. Interactive tools like SpeakBuddy incorporate gamification elements and conversational practice, making the learning process more enjoyable and interactive (Akhila et al., 2024). Results indicate that learners become more confident in their abilities after using these tools, which could be attributed to the immediate feedback and sense of progress they provide (Akhila et al., 2024; Mohammadkarimi, 2024). Furthermore, the ability to practice independently and receive personalized feedback empowers learners, fostering a sense of autonomy in their learning journey (McCrocklin et al., 2022; Shafiee Rad & Roohani, 2024).

In terms of accessibility, ASR-based systems offer unparalleled flexibility in pronunciation learning. As claimed by McCrocklin et al. (2022) and Wang (2024), learners can practice at their own pace, anytime and anywhere, without the constraints of traditional classroom settings. The same tool as the above, SalukiSpeech, provides learners with the option to email summary reports of their practice sessions to instructors, allowing for continuous monitoring and guidance (McCrocklin et al., 2022). This flexibility is particularly beneficial for learners who may have limited access to pronunciation instruction or prefer self-directed learning (McCrocklin et al., 2022; Wang, 2024). ASR technology enables targeted practice by identifying specific areas where learners need improvement. Table 1 shows key features and insights of ASR-based pronunciation learning tools.

Table 1. Key features and insights of ASR-based pronunciation tools.

Authors/Tools	Description	Methodologies	Findings	Limitations
I Love Indonesia (ILI) and NovoLearning (NOVO) (Bashori et al., 2021)	Immediate and corrective feedback on phonetic details	Experimental research; short intervention (5 weeks)	Both ASR systems improved learners’ pronunciation, with NOVO (richer feedback) showing more substantial gains.	Short study duration: no long-term effects measured.
SpeakBuddy (Akhila et al., 2024)	Interactive conversation & instant feedback	Case study	Enhanced learner autonomy and engagement; increased confidence and accuracy in pronunciation.	Focused on one context; generalizability limited.
SalukiSpeech (McCrocklin et al., 2022)	Flexible, student-driven, implicit feedback for sound segments	Tool development; user trials	Segmental accuracy improved, especially with problematic vowels/consonants; learners actively engage in their own process.	Limited practice variety (only one picture description task).
LangGo (Khampusaen et al., 2023)	Contextualized lessons and error feedback aligned with learner backgrounds	Case study: Thai EFL learners	Itemized individual error analysis; improved target segmental features; context improves relevance	Applicability beyond Thai/ESP context not tested.
ASR + Phonetic Analysis System (Lara et al., 2024)	Combination of ASR + phonetic analysis for precision	Technical evaluation with learners	Precise feedback on articulation/intonation, revealing subtle errors hard to catch by human teachers.	Technology-dependent; user interface challenges.
Listnr & Murf (Mohammadkarimi, 2024)	AI tools with instant feedback and consumable practice exercises	Mixed methods	Notable pronunciation gains; increased learner motivation.	Black-box feedback logic; possible over-reliance on technology.
CNN-Based Systems (Rajkumari et al., 2024)	Neural network-powered personalized feedback on accuracy and fluency	Experimental	Personalized, real-time corrections outperformed static approaches; improved both accuracy and fluency.	Resource-intensive; requires large, clean training sets.

This comprehensive analysis (Table 1) highlights the transformative impact of ASR technology on English pronunciation learning, offering personalized, flexible, and engaging solutions for learners. To begin with, LangGo’s contextualized lessons tailored to Thai learners’ linguistic backgrounds improve relevance and target specific phonetic challenges. In a similar vein, NOVO’s rich, immediate corrective feedback accelerates pronunciation gains, enabling learners to make faster, more measurable improvements. Moreover, SalukiSpeech emphasizes learner autonomy and segmental accuracy, addressing problematic vowels and consonants, while ASR + Phonetic Analysis offers precise feedback

on subtle articulation and intonation errors often missed by human teachers. Next, CNN-based systems enhance personalization with real-time corrections for accuracy and fluency, though resource optimization is crucial. SpeakBuddy's interactive approach fosters learner engagement and confidence, showing the importance of active participation. Combining contextualization, real-time feedback, personalization, phonetic precision, and interactive autonomy ensures an effective AI application tailored to Thai learners' needs. While ASR technology has shown considerable promise in facilitating English pronunciation learning across diverse linguistic contexts, a critical research gap persists in its application for Thai learners, particularly within ESP domains. Existing investigations predominantly examine commercially available ASR or AI-based pronunciation tools that are designed for global audiences, rather than purpose-built systems attuned to the unique phonological needs of Thai speakers.

As a result, these tools often provide generic feedback on pronunciation errors, without adequately targeting the parts of the English sound system that are especially problematic for Thais. Adding to the gap is the scarcity of pronounced research on tailoring ASR technology to ESP contexts, where learners must master vocationally or professionally relevant speech patterns in addition to general pronunciation. There is thus a pressing need for the development and empirical evaluation of ASR-driven materials and feedback systems explicitly modeled on situated communicative needs. The present study addresses this gap by investigating whether AI-based instruction can effectively improve Thai learners' pronunciation of English fricative consonants, which represent a particular challenge due to phonological differences between Thai and English.

2.4. Pronunciation Learning in English for Specific Purposes

English for Specific Purposes (ESP) learners face unique challenges related to pronunciation, as they often need to master specialized vocabulary and communicate effectively within specific professional contexts (Basturkmen, 2010). For Thai ESP learners, particularly those in fields such as international business, tourism, engineering, and healthcare, clear pronunciation is essential for professional success (Kaewpet, 2009; Kaya, 2021; Peerachachayanee, 2022).

While existing research has examined the effectiveness of various CAPT and AI applications for pronunciation instruction in different contexts, several gaps remain in the literature. First, studies specifically investigating AI-mediated pronunciation training for Thai learners are limited, despite the unique challenges presented by the Thai phonological system. Second, research focusing on the specific pronunciation needs of ESP learners, particularly regarding fricative consonants crucial for professional English usage, remains underdeveloped. Consequently, the present study addresses these gaps by examining the effectiveness of AI-mediated pronunciation training compared to traditional instruction for Thai ESP learners, with a specific focus on English fricative consonants. By investigating this understudied population and context, this research contributes to our understanding of how AI technology can be leveraged to address specific pronunciation challenges in ESP settings.

In this study, Skills Acquisition Theory (SAT) was applied as the theoretical underpinning for investigating how AI-mediated pronunciation applications can improve English pronunciation among ESP learners. Skills Acquisition Theory posits that skill development progresses through three distinct stages: the cognitive stage (where learners acquire declarative knowledge about pronunciation rules), the associative stage (where knowledge transforms into procedural form through practice and feedback), and the autonomous stage (where pronunciation becomes automatized and requires minimal conscious effort). Consequently, this study employs this theory (SAT) because its structured stages align well

with the process by which AI-mediated applications facilitate progressive improvement in pronunciation skills. It is believed that this application would offer targeted feedback and practice to ESP learners.

As shown in Figure 1, this framework guides the design of the AI application, which provides phonetic instruction and models at the cognitive stage while offering speech recognition and immediate corrective feedback during the associative stage to facilitate the transition toward autonomous pronunciation. The theory helps explain how ESP learners with specific phonological challenges can systematically develop pronunciation skills through AI-mediated practice, addressing segmental features (consonant accuracy). By mapping technological interventions to these established learning stages, this research investigates the effectiveness of AI tools in facilitating the complex process of Thai learners’ pronunciation within specialized English language contexts. By mapping technological interventions to these established learning stages, this research investigates the effectiveness of an AI tool in facilitating the complex process of Thai learners’ pronunciation within specialized English language contexts. Therefore, this study specifically seeks to determine to what extent AI-mediated pronunciation training affects Thai ESP learners’ pronunciation of English fricative consonants.

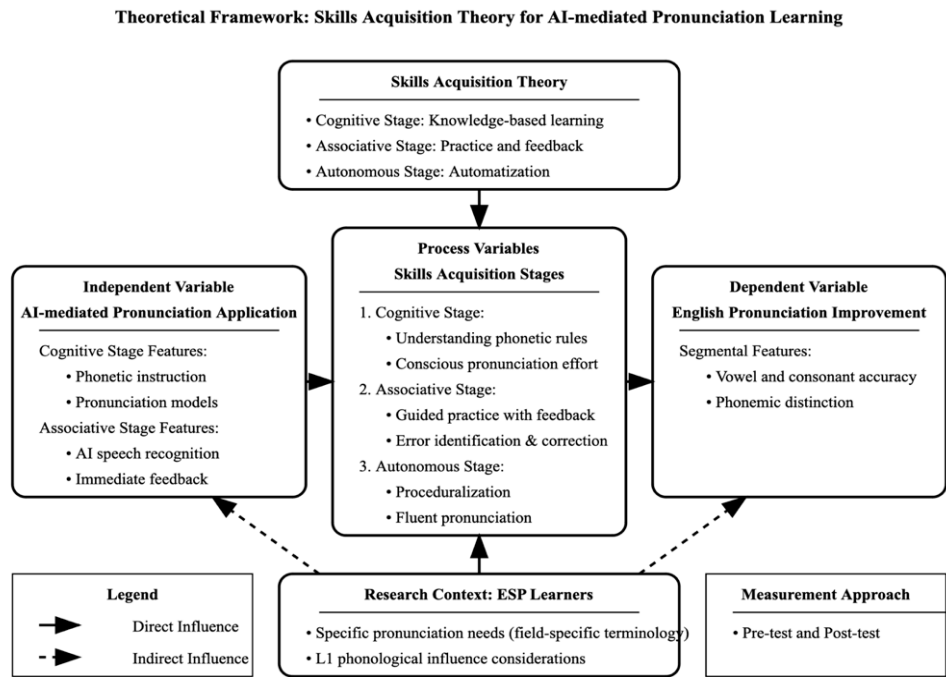


Figure 1. Theoretical framework for AI-mediated pronunciation learning.

Figure 1 illustrates the directional relationships between the theoretical framework’s interconnected components. Initially, the solid arrow from Skills Acquisition Theory to Process Variables shows how the theoretical foundation structures the three-stage learning progression in pronunciation development. Subsequently, the solid arrow from the Independent Variable (AI-mediated Pronunciation Application) to Process Variables indicates how the AI application facilitates learners’ progression through cognitive, associative, and autonomous stages via phonetic instruction, pronunciation models, speech recognition, and immediate feedback.

The solid arrow connecting Process Variables to the Dependent Variable (English Pronunciation Improvement) represents the expected outcome: systematic progression through skills acquisition stages produces measurable improvements in fricative consonant pronunciation accuracy and phonemic distinction. In contrast to these direct relationships,

the dashed arrows from Research Context (ESP Learners) to both the Independent and Dependent Variables signify indirect influences, indicating that the specific characteristics and pronunciation needs of Thai ESP learners, including their L1 phonological system and field-specific terminology requirements, contextually shape both the design features of the AI application and the targeted pronunciation outcomes measured in this investigation.

3. Research Methodology

A quasi-experimental approach was selected due to the inability to assign individual participants to experimental and control conditions within the institutional context. Instead, intact classes were randomly assigned to either the experimental group (receiving AI-mediated pronunciation training) or the control group (receiving traditional pronunciation instruction).

3.1. Participants

The study recruited 74 undergraduate ESP students through convenience sampling from an English for Communication course at a Thai cultural tertiary institution. The participants were assigned to the nonrandomized experimental group ($N = 36$) or control group ($N = 38$) within the available sample. Students represented diverse disciplinary backgrounds, including Thai classical dance and Thai music education, reflecting the varied professional contexts characteristic of ESP instruction in Thailand’s tertiary education sector. Participants ranged in age from 19 to 22 years and had studied English as a Foreign Language (EFL) for an average of 12 years through Thailand’s formal education system. Prior to the intervention, all participants completed a background questionnaire to gather information about their English learning history, previous exposure to pronunciation instruction, and familiarity with technology-enhanced language learning. Statistical analyses were conducted using SPSS Statistic version 29 from IBM (Armonk, NY, USA). To ensure that the experimental and control groups were comparable in terms of initial pronunciation ability, a pre-test was administered to all participants. Statistical analysis confirmed that there were no significant differences between the groups at baseline $t(72) = -0.73, p = 0.468$), establishing a sound foundation for subsequent comparisons. Table 2 demonstrates the results of test of normality for the control and experimental groups.

Table 2. Test of normality within control and experimental groups.

Group		Kolmogorov–Smirnov			Shapiro–Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Pretest	Exp	0.12	38	0.174	0.96	38	0.186
	Control	0.14	36	0.077	0.96	36	0.216
Posttest	Exp	0.11	38	0.200	0.98	38	0.579
	Control	0.09	36	0.200	0.97	36	0.540

Prior to the main analyses, tests of normality using Kolmogorov–Smirnov and Shapiro–Wilk were performed to verify that the data met the assumptions for parametric statistical tests. Both tests confirmed normal distribution of the pronunciation scores across conditions ($p > 0.05$).

3.2. Research Instruments

3.2.1. English Pronunciation Test

The English pronunciation test was administered as a read-aloud evaluation designed to evaluate participants’ production accuracy of nine target fricative consonants (/f/, /v/, /θ/, /ð/, /s/, /z/, /ʃ/, /ʒ/, and /h/) in various phonetic positions (see Appendix A).

Each fricative was systematically tested in initial, medial, and final positions where phonologically permissible in English phonology. The test consisted of 60 individual words and 25 sentences containing multiple target sounds, selected to ESP-specific vocabulary relevant to the participants' fields of study (e.g., Khon masked play, Thai music instruments, Thai dance, and drama), tailoring training and learning opportunities to the specific needs and professional realities of the learners (Lao-un & Bunyaphithak, 2025). The test was administered individually to each participant in a controlled laboratory setting. The example of the sentences targeting fricatives in ESP-relevant contexts sounds in each position as shown:

- The folk theatre performance requires careful choreography.
- Traditional musicians use string instruments like the zither.
- They preserve ancient methods of dance.

The sentences follow similar patterns, each targeting 3–5 fricatives in professionally relevant contexts. The target words and sentences were presented in randomized order to prevent participants from identifying patterns in target words, thereby enhancing the validity of the assessment. To establish content validity, the words and sentences were reviewed by three experts in the field of English with teaching backgrounds in phonetics and pronunciation pedagogy. Modifications were implemented based on their expert feedback to ensure linguistic appropriateness and proper representation of the target phonemes. Content validity was established through expert review, with items revised based on feedback to ensure appropriate representation of the target sounds and contexts, resulting in Intraclass Correlation Coefficient (ICC) value of 0.86, indicating the agreement or reliability among multiple experts scoring a pronunciation test.

Prior to actual implementation, a pilot test was conducted with 12 stratified-selected participants representing different proficiency levels (high and low English ability), which represented participants from both experimental and control group contexts. This piloting procedure followed identical administration protocols planned for the main study to identify potential procedural issues and estimate completion time. Results from the pilot indicated the task was accessible to learners across proficiency levels, with an average completion time of 30 min, confirming the instrument's feasibility for the intended population.

3.2.2. AI-Mediated Pronunciation Application

The experimental group used a specifically designed program for Thai ESP learners of English. The AI application lessons were tailored to the phonological challenges faced by Thai speakers, contextualized for Thai context (Kanokpermpoon, 2007; Naruemon et al., 2024; Peerachachayane, 2022). The development of the AI-mediated pronunciation learning application was guided by the theoretical foundation of Skill Acquisition Theory (SAT) (R. DeKeyser, 2007b). The primary objective of the application is to provide personalized, context-specific pronunciation training that supports learners through the cognitive, associative, and autonomous stages of skill development.

The application was collaboratively developed by the lead researcher, who designed the pedagogical framework, selected linguistic content, and conducted the learner needs analysis, and a programmer, who was responsible for implementing the user interface, integrating the automatic speech recognition (ASR) engine, and optimizing system feedback algorithms. The following diagram (Figure 2) indicates the architecture of automatic speech recognition of this current study.

Sounds to Speak Application Architecture

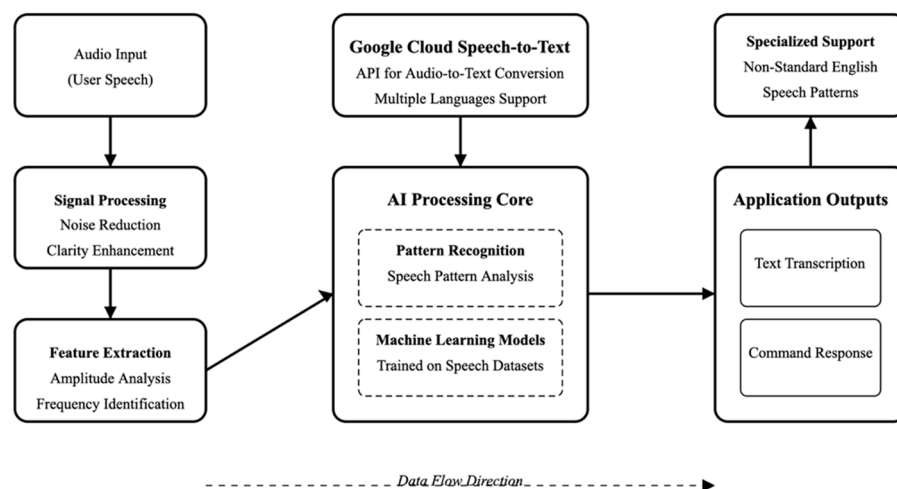


Figure 2. The architecture of AI-based automatic speech recognition of this study.

In this study, AI speech recognition applications involve using artificial intelligence to convert spoken words into text or to understand and respond to spoken commands. AI speech recognition systems use complex algorithms and machine learning models to convert spoken words into text. The process typically involves several steps: (1) Signal Processing: The audio input is processed to reduce noise and enhance clarity. (2) Feature Extraction: Key characteristics of the speech signal, such as amplitude and frequency, are identified. (3) Pattern Recognition: AI systems analyze patterns in the audio to identify words and phrases. (4) Machine Learning Model. These models are trained on large datasets of speech and text to learn the relationships between sounds and words. AI speech recognition utilizes the Google Cloud Speech-to-Text API to convert audio to text with support for multiple languages and various customization options. The AI-powered pronunciation training application is specifically designed for non-native English speakers, incorporating automatic speech recognition technology that is calibrated to detect and provide feedback on pronunciation errors commonly made by L2 English learners.

AI speech recognition uses API, and the tool is Google Cloud Speech-to-Text, which is a powerful API for converting audio to text, offering various languages and customization options. The AI-powered pronunciation training application incorporates automatic speech recognition technology calibrated to detect and provide feedback on pronunciation errors common among L2 English learners (Neri et al., 2003). Following the ASR-based CAPT system, key features of the application included the following:

1. Speech analysis capability that identifies specific errors in fricative production;
2. Explicit and implicit textual feedback on specific pronunciation errors;
3. Personalized practice exercises targeting individual learners' difficulties;
4. Progress tracking and analytics;
5. Specialized vocabulary modules relevant to participants' ESP contexts.

The application was selected based on a comprehensive evaluation of available technologies, with particular attention to its capacity to address the specific pronunciation challenges faced by Thai learners of English. Prior to implementation, the application's speech recognition accuracy was tested with a sample of Thai speakers to ensure its appropriateness for the target population. This application was available only on iOS devices. The interface employs a systematic color-coding scheme to facilitate learning: green backgrounds and buttons indicate correct pronunciation responses, red backgrounds signal incorrect attempts with specific error identification, and purple elements denote

navigational functions and the competitive scoreboard feature. The microphone symbol prompts learners to record their pronunciation, while speaker icons provide access to English speaker models. Figure 3 demonstrates the overview of the application's features.

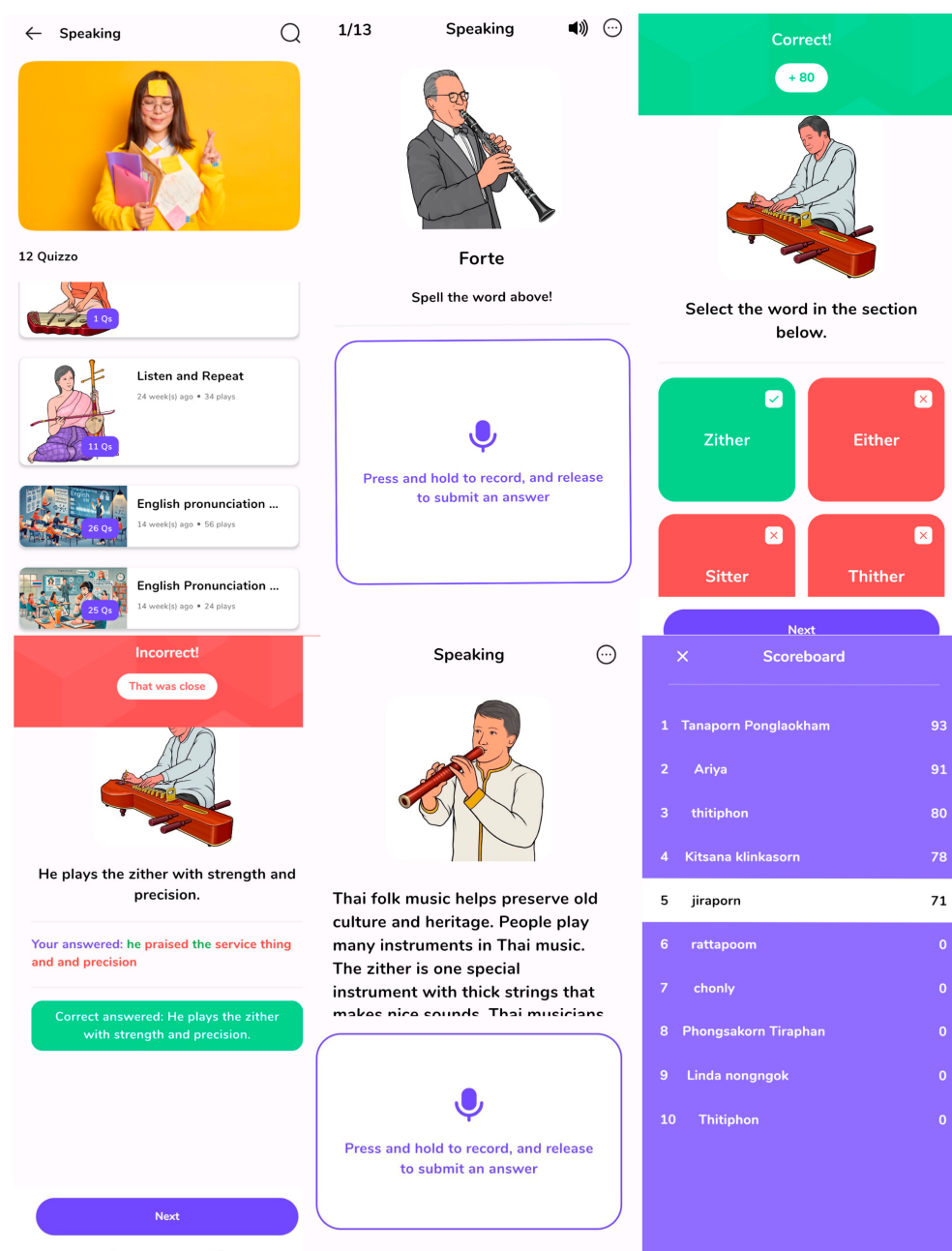


Figure 3. The overview of the Sounds to Speak features.

The features of the application were systematically designed to correspond to each stage of SAT: in the cognitive stage, learners are introduced to phonetic rules and pronunciation models through auditory input; in the associative stage, ASR technology delivers immediate feedback on learners' spoken output, allowing real-time identification and correction of pronunciation errors; and in the autonomous stage, learners engage in extended, ESP contextualized pronunciation drills, focusing on Thai dance and classical music, to promote proceduralization and fluency (Kachinske, 2021). In addition, the application's features, including minimal pair, real-time speech input, and a dynamic scoreboard, foster motivation and learner autonomy, while its feedback mechanisms support accurate segmen-

tal production, particularly of English consonants commonly problematic for Thai speakers. In the valuation phase, a pilot test with 10 target users informed iterative refinements including ASR recalibration and interface adjustments. The ASR recalibration process was completed by the programmer and researchers through systematic iterative testing with representative target users to optimize accuracy for Thai-accented English. Following initial development using Google Cloud Speech-to-Text API, a comprehensive pilot study with 10 Thai speakers representing the target demographic informed critical refinements to the speech recognition algorithms. The calibration process involved adjusting sensitivity thresholds specifically for Thai-accented English pronunciation patterns, with particular attention to common substitution errors identified in preliminary testing. The final version of the application demonstrated strong usability, user satisfaction, and pedagogical alignment with the Thai ESP context.

3.3. Research Procedures

Prior to data collection, the study received ethical approval from the researcher's institution. All participants provided written informed consent after receiving comprehensive information about study objectives, procedures, potential benefits and risks, and their rights as research participants. The study was conducted over a 16-week period during a regular academic semester. Both groups received 40 h of pronunciation instruction (4 h per week), with the experimental group using the AI-mediated pronunciation application and the control group receiving traditional pronunciation instruction. The researcher maintained instructional consistency by teaching both experimental and control groups using identical materials and lesson plans, thus controlling for teacher effects. Following Skill Acquisition Theory's three-phase developmental progression ([Anderson, 1982](#); [R. M. DeKeyser, 2015](#)). The instructional sequence systematically guided learners from the cognitive stage through the associative stage to the autonomous stage of skill development. Each weekly instructional cycle began with face-to-face sessions focusing on the cognitive stage of skill acquisition, during which participants in both groups received explicit declarative knowledge about phonological features through minimal-pair discrimination tasks, explicit articulatory instruction, and metalinguistic awareness activities. Following this cognitive foundation, the experimental group transitioned to AI-mediated practice sessions where they engaged with the pronunciation application designed to provide individualized feedback and adaptive scaffolding to support the transition to the associative stage of acquisition.

3.3.1. Pre-Testing Phase

Prior to the intervention, all participants completed the English fricative pronunciation test, which was recorded for subsequent analysis. Each participant was tested individually in a quiet room to ensure optimal recording quality.

3.3.2. Intervention Phase

Participants in the experimental group received comprehensive training on the AI-mediated pronunciation application before beginning the intervention sessions. Each experimental session followed a systematic format beginning with a 30-min teacher-led introduction to target fricative sounds, establishing foundational understanding of articulatory features and phonetic characteristics. This introductory phase was followed by 90 min of individualized practice using the AI application, during which participants engaged with real-time feedback mechanisms, completed targeted pronunciation exercises, and monitored their progress through the application's tracking features. Throughout the individual practice sessions, the instructor maintained an active supervisory role, circulating

among participants to provide technical support, troubleshoot application issues, and offer supplementary guidance when required.

The control group received traditional pronunciation instruction targeting identical fricative sounds through established pedagogical approaches. Control group sessions maintained the same temporal structure with a 30-min teacher-led introduction to target sounds, followed by 90 min of conventional pronunciation activities. These activities encompassed listen-and-repeat exercises designed to develop sound recognition and production skills, minimal pair discrimination tasks to enhance phonemic awareness, guided practice sessions for controlled production, and structured peer feedback opportunities. The instructor delivered feedback and correction using proven techniques including explicit modeling of correct articulation, detailed explanations of articulatory processes, and corrective feedback during practice activities. Both experimental and control groups followed an identical phonetic curriculum structured around a progressive sequence of fricative sound introduction. The following Figure 4 demonstrates the teaching methods of the study:

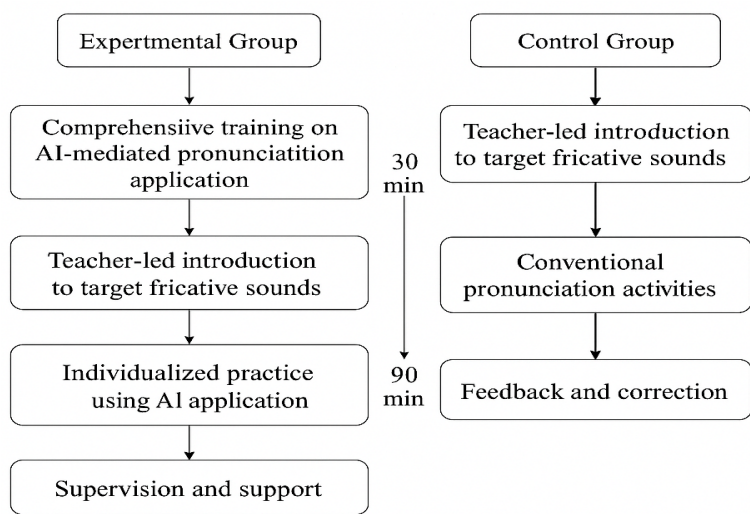


Figure 4. Teaching methods of experimental and control groups.

The syllabus strategically began with fricative sounds that have Thai language equivalents, specifically /f/, /s/, and /h/, allowing participants to build confidence with familiar articulatory patterns before progressing to more challenging sounds. Subsequently, instruction advanced to fricative sounds without Thai counterparts, including /v/, /θ/, /ð/, /z/, /ʃ/, and /ʒ/, which represent particular pronunciation challenges for Thai learners. To maintain relevance to participants’ professional contexts, both groups incorporated English for Specific Purposes vocabulary directly related to their fields of study in dramatic arts, ensuring that pronunciation practice occurred within meaningful and professionally relevant contexts. The pedagogical framework for both interventions was grounded in Skill Acquisition Theory, which structured the learning progression from initial declarative knowledge acquisition during the cognitive stage, through proceduralization during the associative stage, and ultimately toward automaticity achievement in the autonomous stage. Table 3 indicates the instructional plan for English pronunciation teaching between the control and experimental groups.

Table 3. Instructional plan for control and experimental groups based on Skill Acquisition Theory.

Week	Instructional Stage (SAT)	Focus Content	Control Group (Traditional Instruction)	Experimental Group (AI-Mediated Practice)
1–2	Cognitive (Declarative)	- Pre-Instruction Evaluation - Introduction to English fricatives - Sounds with Thai equivalents (/f/, /s/, /h/) - Sounds without Thai counterparts (/v/, /θ/, /ð/, /z/, /ʃ/, /ʒ/) - Sounds without Thai counterparts (/v/, /θ/, /ð/, /z/, /ʃ/, /ʒ/)	- Teacher-led articulatory instruction with visual aids - Metalinguistic awareness activities - Explicit articulatory instruction - Auditory identification exercises - Minimal-pair discrimination tasks - Guided sentence drills - Teacher-assisted feedback	- Same content delivery with support from AI-driven multimodal input (audio and visual) and corrective feedback - Metalinguistic awareness activities - Explicit articulatory instruction - AI-driven multimodal input (Minimal-pair discrimination tasks and Guided sentence drills)
3–4	Cognitive to Associative	- Sounds without Thai counterparts (/v/, /θ/, /ð/, /z/, /ʃ/, /ʒ/)	- Teacher-assisted feedback	- AI-driven multimodal input (Minimal-pair discrimination tasks and Guided sentence drills)
5–6	Associative (Procedural)	Voiced fricatives (/v/, /z/)	- Explicit articulatory instruction - Targeted word and sentence practice with profession-related context - Teacher scaffolding with controlled drills and contextual examples	- Explicit articulatory instruction - Targeted word and sentence practice with profession-related context - Teacher scaffolding with controlled drills and contextual examples
7–8	Associative	Interdental fricatives (/θ/, /ð/)	- Mini-presentations for teacher feedback	- Personalized and contextualized practice exercises via AI
9–10	Associative to Autonomous	Post-alveolar fricatives (/ʃ/, /ʒ/)		- Real-time error detection and feedback via AI - Mini-presentations for teacher feedback - Scenario-based AI tasks - Contextual pronunciation simulations
11–12	Autonomous	Sentence-level and ESP wordlist practice	- ESP-specific vocabulary and sentence practice - Teacher feedback	- AI-driven multimodal input (audio and visual) - Teacher feedback integrated into AI review
13–14	Autonomous	Contextualized speech production	- ESP-specific vocabulary practice - Controlled pronunciation task (oral presentation) - Teacher feedback	- Controlled pronunciation task (oral presentation) - AI-driven multimodal input (audio and visual) - Teacher feedback integrated into AI review
15–16	Post-Instruction Evaluation	Pronunciation test (ESP-specific words) and reflective evaluation	- Post-test - AI analyzes progress from pre- to post-test	- Same post-test - AI analyzes progress from pre- to post-test - Reflective questionnaire

3.3.3. Post-Testing Phase

Following the 16-week intervention, all participants completed the post-test using the same pronunciation test and procedures as in the pre-test. The recordings were evaluated by the same panel of raters, who were blind to group assignment and whether each recording was from the pre-test or post-test to minimize potential rating bias.

3.4. Data Analysis

The pronunciation test recordings via the AI-based application were evaluated using accuracy evaluation adapted from [Khampusaen et al. \(2023\)](#) for each target sound, with scores aggregated to provide a pronunciation accuracy rate of learners’ pronunciation performance (Table 4). The provided scores are from 1 to 8, where 1 indicates less than 30 percent accuracy of pronunciation.

Table 4. AI-Based pronunciation assessment scoring rubric.

AI Accuracy	Rating Scale	Descriptions
>90%	8	The rater would perceive the fricative as consistently and clearly produced, closely matching native-like pronunciation with minimal or no discernible errors.
80–90%	7	
70–80%	6	The rater might notice some inconsistencies or minor deviations in the production of the fricative, but the sound remains largely intelligible.
60–70%	5	
50–60%	4	
40–50%	3	The rater would identify significant and frequent errors in the production of the fricative, potentially leading to reduced intelligibility
30–40%	2	
<30%	1	

The speech samples from both control and experimental groups were analyzed through a systematic approach to ensure measurement reliability. To establish measurement consistency between automated and human assessment methods, the researcher implemented a stratified sampling approach, randomly selecting five representative recordings from each group (5 of 36 from the experimental group and 5 of 38 from the control group) for dual evaluation by both the AI-based application and two experienced English language teaching researchers/teachers who were not native speakers of the language. The five-sample selection from each group provides adequate statistical power for correlation analysis between automated and human assessment methods while representing approximately 13–14% of each group’s total recordings due to resource constraints associated with comprehensive human rating of the entire dataset. In addition, the random selection process ensures that the chosen samples maintain representativeness across the range of participant performance levels within each group.

In the data analysis for human raters, the raters were not informed which samples belonged to the control or experimental groups, nor whether they were from pre- or post-test conditions, reducing potential expectation bias. They implicitly assign a qualitative accuracy level to the individual sound based on their expert judgment. This qualitative assessment then corresponds to the percentage ranges outlined in your rubric, allowing them to assign the appropriate numerical rating (1–8) for that specific fricative’s pronunciation. For instance, if a fricative is consistently produced with only minor, occasional inaccuracies that do not hinder intelligibility, a human rater might deem its accuracy to be in the 80–90% range, resulting in a rating of 7.

Following the inter-rater reliability establishment, the remaining pre-test and post-test were transcribed by the AI-based application. Paired-samples *t*-tests were conducted to examine within-group improvements from pre-test to post-test for both experimental and control groups. Independent-sample *t*-tests were employed to compare between-group differences in both overall pronunciation scores and specific fricative sound performance. Effect sizes were calculated to determine the magnitude of observed differences, with values of 0.4, 0.7, and 1.00 considered small, medium, and large effects, respectively (Plonsky & Oswald, 2014).

4. Results

To address the research question regarding the extent to which AI-mediated pronunciation affects Thai ESP learners’ pronunciation ability, this section presents the quantitative analysis of the pronunciation test results. The analysis employed parametric tests to compare both within-group improvements (before and after intervention) and between-group differences (AI-based instruction versus traditional methods). Statistical analyses focused on overall pronunciation performance and specific performance on nine English fricative sounds.

4.1. Within-Group Comparisons: Pre-Test and Post-Test Performance

A paired-samples *t*-tests were conducted to examine the effectiveness of both instructional approaches by comparing pre-test and post-test scores within each group.

The results presented in Table 5 reveal significant improvements in pronunciation performance for both groups. The control group, who received traditional pronunciation instruction, demonstrated a statistically significant improvement from pre-test (*M* = 85.45, *SD* = 15.39) to post-test (*M* = 121.16, *SD* = 23.98), *t*(37) = −15.10, *p* < 0.001, which represented a larger effect size (*d* = −2.45). Similarly, the experimental group, which received AI-based pronunciation training, showed a significant improvement from pre-test (*M* = 87.67, *SD* = 10.07) to post-test (*M* = 135.03, *SD* = 16.99), *t*(35) = −14.21, *p* < 0.001, also with a large effect size (*d* = 2.32). These results indicate that while both instructional methods led to significant and substantial improvements in pronunciation, the gains were particularly notable in both conditions.

Table 5. Paired-Samples *t*-Test results of pre-test and post-test scores.

Test	Group	<i>M</i>	<i>SD</i>	MD	<i>df</i>	<i>t</i>	<i>p</i>	Cohen’s <i>d</i>
Pretest	Exp (<i>N</i> = 36)	87.67	10.07	47.36	35	−14.21	<0.001	2.32
Posttest		135.03	16.99					
Pretest	Control (<i>N</i> = 38)	85.45	15.39	35.71	37	−15.10	<0.001	−2.45
Posttest		121.16	23.98					

Note. *M* = Mean, *SD* = Standard Deviation, MD = Mean Difference, Cohen’s *d* represents effect size.

4.2. Between-Group Comparisons: Overall Pronunciation Performance

An independent samples *t*-test was conducted to compare pretest scores between the control and experimental groups. The results are presented in Table 6.

The control group, which received traditional pronunciation instruction (*M* = 85.45, *SD* = 15.39), and the experimental group, which received AI-mediated pronunciation learning (*M* = 87.67, *SD* = 10.07), did not differ significantly, *t*(72) = −0.73, *p* = 0.468. The 95% confidence interval of the difference between means ranged from −8.28 to 3.84. The effect size was small (*d* = −0.17), indicating that the two groups started from comparable baseline levels.

Table 6. Independent-Sample *t*-Test results comparing overall pronunciation scores.

Test	Group	<i>M</i>	<i>SD</i>	<i>df</i>	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
Pre-test	Control (<i>N</i> = 38)	87.67	15.39	72	0.73	0.468	−0.17
	Exp (<i>N</i> = 36)	85.45	16.99				
Post-test	Control (<i>N</i> = 38)	121.16	23.98	72	2.86	0.006	−0.66
	Exp (<i>N</i> = 36)	135.03	16.99				

Note. *M* = Mean, *SD* = Standard Deviation, *t* = *t*-test statistic, *df* = degrees of freedom, *p* = *p*-value. Cohen's *d* represents effect size.

After the intervention, both groups showed significant improvements in pronunciation performance following their respective instructional approaches. In the posttest, the experimental group (*M* = 135.03, *SD* = 16.99) scored significantly higher than the control group (*M* = 121.16, *SD* = 23.98), $t(72) = -2.86$, $p = 0.006$. The 95% confidence interval of the difference between means ranged from −23.55 to −4.19. The effect size was moderate ($d = -0.66$). This result indicates a medium-sized effect, suggesting that the experimental intervention had a meaningful impact on learners' performance. At the group level, although considerable individual variation was observed, the greatest amount of variation (*SD* = 23.98) occurred in the control group's posttest results. This indicates that not all participants responded equally well to traditional instruction, which may have contributed to the wider confidence intervals observed in the effect size estimates.

4.3. Between-Group Comparisons: Specific Fricative Sounds

To provide a clearer understanding of the differential effects of AI-based instruction versus traditional methods, independent-sample *t*-tests were conducted to compare the post-test performance on nine specific English fricative sounds.

The results presented in Table 7 reveal a pattern of differential improvement across the nine fricative sounds. Statistically significant differences between the experimental and control groups were observed for five fricative sounds: /θ/, /ð/, /z/, /ʃ/, and /h/. For the interdental fricative, the experimental group (*M* = 15.81, *SD* = 3.85) significantly outperformed the control group (*M* = 13.79, *SD* = 3.76), $t(72) = -2.28$, $p = 0.026$, with a moderate effect size ($d = -0.53$). Similarly, for the voiced interdental fricative /ð/, the experimental group (*M* = 15.06, *SD* = 3.76) demonstrated superior performance compared to the control group (*M* = 13.32, *SD* = 3.30), $t(72) = -2.12$, $p = 0.037$, with a moderate effect size ($d = -0.49$).

The most pronounced differences were observed in the pronunciation of the alveolar fricative /z/, where the experimental group (*M* = 17.22, *SD* = 2.68) significantly outperformed the control group (*M* = 14.71, *SD* = 4.45), $t(72) = -2.92$, $p = 0.005$, with a moderate-to-large effect size ($d = -0.68$). Similarly, for the post-alveolar fricative /ʃ/, the experimental group (*M* = 17.75, *SD* = 2.61) demonstrated markedly better performance than the control group (*M* = 15.39, *SD* = 4.48), $t(72) = -2.74$, $p = 0.008$, with a moderate-to-large effect size ($d = -0.64$). Additionally, the glottal fricative /h/ showed a significant difference favoring the experimental group (*M* = 11.75, *SD* = 2.60) over the control group (*M* = 10.29, *SD* = 2.20), $t(72) = -2.61$, $p = 0.011$, with a moderate effect size ($d = -0.61$). However, no statistically significant differences were observed between the groups for the remaining four fricatives /f/ and /s/, the voiced fricative /v/, and the post-alveolar fricative /ʒ/ ($p > 0.05$). Although the experimental group consistently achieved higher mean scores for these sounds, the differences did not reach statistical significance.

Table 7. Independent-Sample *t*-Test results comparing specific fricative sounds.

Fricative	Group	M	SD	df	<i>t</i>	<i>p</i>	Mean Difference	Cohen's <i>d</i>
/f/	Control (N = 38)	14.32	2.27	72	−1.89	0.063	−1.02	−0.44
	Exp (N = 36)	15.32	3.59					
/v/	Control (N = 38)	15.18	3.59	72	−0.86	0.394	−0.076	−0.20
	Exp (N = 36)	15.94	4.04					
/θ/	Control (N = 38)	13.79	3.76	72	−2.28	0.026 *	−2.02	−0.53
	Exp (N = 36)	15.81	3.85					
/ð/	Control (N = 38)	13.32	3.30	72	−2.12	0.037 *	−1.74	−0.49
	Exp (N = 36)	15.06	3.76					
/s/	Control (N = 38)	14.76	7.88	72	−1.37	0.175	−1.90	−0.32
	Exp (N = 36)	16.67	2.82					
/z/	Control (N = 38)	14.71	4.45	72	−2.92	0.005 **	−2.51	−0.68
	Exp (N = 36)	17.22	2.68					
/ʃ/	Control (N = 38)	15.39	4.48	72	−2.74	0.008 **	−2.36	−0.64
	Exp (N = 36)	17.75	2.61					
/ʒ/	Control (N = 38)	9.39	3.12	72	−0.16	0.877	−0.11	−0.04
	Exp (N = 36)	9.50	2.67					
/h/	Control (N = 38)	10.29	2.20	72	−2.61	0.011 *	−1.46	−0.61
	Exp (N = 36)	11.75	2.60					

Note * $p < 0.05$. ** $p < 0.01$, *M* = Mean, *SD* = Standard Deviation, Cohen's *d* represents the effect size.

5. Discussion

The present study offers a comprehensive evaluation of AI-mediated pronunciation training compared to traditional methods for Thai ESP learners, specifically in the acquisition of English fricative consonants. The results demonstrate that while both instructional approaches yield significant pronunciation gains, participants utilizing the AI application experienced substantially greater improvements. To state the difference, the findings of the research were corroborated by the data obtained from paired-samples *t*-tests, independent-sample comparisons, and independent-sample *t*-test comparing specific fricative sounds, presenting a coherent picture of AI-mediated instruction's advantages. While both groups demonstrated significant within-group improvements ($p < 0.001$ for both experimental and control groups), the experimental group's substantially moderate effect size ($d = 0.17$, $d = 0.66$) and mean difference (mean difference of 47.36 and 35.71 in the control group) suggest that AI technology facilitates more efficient skill acquisition than traditional methods

alone. This pattern becomes more pronounced when examining between-group comparisons. The absence of significant pre-test differences ($p = 0.468$) followed by significant post-test advantages for the experimental group ($p = 0.006$, $d = -0.66$) demonstrates that the significant improvements resulted from instructional differences rather than initial ability variations. Further, it was determined from between-group comparisons in specific fricative sounds that AI-based pronunciation training was particularly effective in enhancing the pronunciation of specific fricative sounds that are typically challenging for Thai ESP learners, especially the interdental fricatives (/θ/ and /ð/), which do not exist in the Thai phonological system, as well as the sibilants /z/ and /ʃ/, and the glottal fricative /h/. Nevertheless, the absence of significant differences for the four fricatives /f/, /s/, /v/, /ʒ/ can be attributed to several interconnected phonological and pedagogical factors. This suggests that integrating advanced technology into pronunciation pedagogy can more effectively target persistent segmental pronunciation errors, particularly those stemming from L1–L2 phonological differences.

Building upon these findings, the segment-specific patterns observed in this study provide compelling support for theoretical predictions derived from both the Speech Learning Model (Flege & Bohn, 2021) and the Perceptual Assimilation Model of Second Language Speech Learning (Best & Tyler, 2007). Importantly, the differential effectiveness of AI-mediated instruction across fricative categories aligns closely with SLM-r’s fundamental principle that cross-linguistic phonetic similarity determines acquisition difficulty. According to SLM-r’s Perceived Dissimilarity Principle, sounds that are perceived as phonetically dissimilar to any existing L1 category are more likely to result in successful new phonetic category formation, whereas sounds perceived as similar to L1 categories resist new category establishment. In this context, the significant improvements observed for interdental fricatives /θ/ and /ð/ exemplify this principle, as these sounds have no Thai equivalents and thus represent what SLM-r terms “new” sounds that can readily form distinct phonetic categories. In contrast, the lack of significant improvement for /v/, despite its absence from Thai, may reflect PAM-L2’s Single Category assimilation pattern, where /v/ consistently assimilates to Thai /w/ or /f/ (Kanokpermpoon, 2007; Kitikanan, 2017; Isarankura, 2016).

The results can be interpreted through the lens of Skill Acquisition Theory (SAT), which posits that language skills develop through cognitive, associative, and autonomous stages. However, integrating perceptual phonology frameworks provides a more comprehensive explanation for the differential learning outcomes observed across fricative sounds. Additionally, the AI application appears to have accelerated the transition from declarative to procedural knowledge by providing consistent, immediate feedback critical for the procedural stage. The AI application appears to have facilitated more efficient transition between these stages compared to traditional instruction (Bashori et al., 2021; McCrocklin et al., 2022; Mohammadkarimi, 2024). This efficiency may result from the application’s ability to provide both explicit articulatory instruction and implicit perceptual training through varied input (Strange, 2011). During the cognitive stage, both groups received explicit instruction on fricative articulation; however, during the associative stage, the AI application provided immediate, individualized feedback and repetitive practice that allowed learners to proceduralize their declarative knowledge more effectively (R. M. DeKeyser, 2009; Taie, 2014). Consequently, this finding challenges the assumption that all pronunciation features are equally difficult to acquire and supports the theoretical position that L1–L2 phonological distance is a critical factor in determining learning outcomes (Huang & Kitikanan, 2022; Kitikanan, 2017; Naruemon et al., 2024). Furthermore, the pronounced advantages observed for /z/ and /ʃ/ can be understood through PAM-L2’s assimilation taxonomy. Thai learners typically exhibit Two-Category assimilation for the /s/-/z/ contrast, where both sounds assimilate to Thai /s/, creating discrimination diffi-

culties that persist in traditional instruction (Kitikanan, 2017; Naruemon et al., 2024). The AI application's capacity to provide systematic acoustic contrasts and immediate feedback appears to have enhanced perceptual sensitivity to the voicing distinction, facilitating what SLM-r terms "category splitting", which is a process whereby learners dissimilate previously merged L1–L2 categories to establish phonetic contrast (Kanokpermpoon, 2007; Kitikanan, 2017; Naruemon et al., 2024).

These results align with previous research by McCrocklin (2019) and Khampusaen et al. (2023), who found that AI-powered applications can effectively enhance L2 pronunciation through personalized feedback and targeted practice. The current study extends these findings specifically to Thai ESP learners, a population that faces unique challenges due to the significant phonological differences between Thai and English (Naruemon et al., 2024). The significant advantage observed in the experimental group may be attributed to several features of AI-mediated instruction that address specific needs in pronunciation learning (Khampusaen et al., 2023). First, the immediate and consistent feedback provided by the AI application allows learners to make real-time adjustments to their articulation patterns, consistent with R. Schmidt's (1990) Noticing Hypothesis, which emphasizes the importance of conscious attention to form in language acquisition.

Traditional instruction, while effective, cannot provide the same level of individualized feedback to all learners simultaneously, potentially limiting opportunities for noticing and correction. Additionally, AI-mediated instruction allowed learners to focus more intensively on their individual difficulties rather than following a one-size-fits-all approach. As noted by Thomson and Derwing (2015), L2 pronunciation challenges are often specific to a learner's L1 background and individual characteristics, which affect perceived phonetic difficulty according to the language-specific filter through which novel language input is processed (Flege & Bohn, 2021). The adaptive algorithms in AI applications can identify and target these specific challenges, potentially accelerating the acquisition process.

5.1. Differential Effects on Specific Fricative Sounds

The quantitative results reveal three distinct patterns of fricative acquisition under AI-mediated versus traditional English pronunciation instruction, each reflecting different aspects of cross-linguistic phonological transfer and technological intervention effectiveness. A particularly noteworthy finding of this study is the differential impact of AI-mediated instruction on specific fricative sounds. The experimental group significantly outperformed the control group on five fricatives (/θ/, /ð/, /z/, /ʃ/, and /h/), while no significant differences were observed for the remaining four fricatives (/f/, /v/, /s/, and /ʒ/). The findings are consistent with the findings of Kanokpermpoon (2007), Isarankura (2016) and Kitikanan (2017) that the inexistence of English fricatives in Thai imposes pronunciation challenges for Thai learners.

These results confirm that skill acquisition is highly specific, with limited transfer between skills (R. M. DeKeyser, 2015). The sounds that showed the greatest improvement under AI-mediated instruction include those that represent the most significant cross-linguistic challenges for Thai learners. The interdental fricatives (/θ/ and /ð/) have no equivalent in Thai and are typically substituted with stops (/t/ and /d/) or other fricatives (/s/ and /z/) (Naruemon et al., 2024), thereby making them particularly difficult to acquire (Kanokpermpoon, 2007). Similarly, the voiced alveolar fricative /z/ is often confused with its voiceless counterpart /s/ by Thai speakers due to the lack of voiced-voiceless fricative contrasts in Thai (Isarankura, 2016). This finding would align with Flege's (1995) SLM, in which phones perceived as "similar" due to cross-linguistic interaction are also most difficult.

The AI application's ability to detect these specific substitution patterns, provide repetitive practice, and targeted feedback appears to have been particularly beneficial for these challenging sounds. In AI-based pronunciation training, repetition involves learners repeatedly listening to and imitating target sounds or phrases to improve their accuracy and fluency, which can lead to increased familiarity and mastery of the target sounds (Dennis, 2024). In contrast, sounds like /f/ and /s/, which have closer equivalents in Thai, showed improvement under both instructional approaches, suggesting that traditional methods may be sufficient for sounds that are less cross-linguistically challenging. The significant improvement in /h/ pronunciation was unexpected, as this sound exists in Thai and typically presents fewer difficulties for Thai learners. This finding may be related to the specific contexts in which /h/ was tested, possibly including word-final positions or consonant clusters that do not occur in Thai. New sounds that are perceived as phonetically dissimilar to any existing L1 category are more likely to result in the formation of a new, distinct phonetic category, facilitating their acquisition (Flege & Bohn, 2021). Alternatively, it could reflect the AI application's sensitivity to subtle aspects of articulation that might not be addressed in traditional instruction.

However, the lack of significant difference for /ʒ/, despite its absence in the Thai phonological system, may be attributed to the relatively low frequency of this sound in English and consequently less practice with it during the intervention period (Isarankura, 2016). As was found in Isarankura (2016) and Kitikanan (2017) that sounds that do not exist in the Thai phonological inventory are particularly challenging for Thai learners to both perceive and produce. This suggests that even with AI-mediated instruction, sufficient exposure and practice remain essential factors in pronunciation acquisition.

5.2. Implications for ESP Pronunciation Pedagogy

The findings of this study have important implications for pronunciation pedagogy in ESP contexts. First, they suggest that AI-mediated pronunciation training can be a valuable complement to traditional instruction, particularly for addressing specific phonological challenges faced by learners from particular L1 backgrounds. For Thai ESP programs, incorporating AI applications that target problematic fricatives could enhance learners' communicative effectiveness in professional contexts.

The differential effectiveness observed across fricative sounds suggests that a hybrid approach combining AI-mediated and traditional instruction might be optimal. While AI applications excel at providing personalized feedback on particularly challenging sounds, traditional instruction may be equally effective for sounds that present fewer cross-linguistic difficulties. The success of AI-mediated pronunciation training in this study suggests potential applications for other aspects of ESP instruction. Similar approaches could be developed for discipline-specific vocabulary and discourse patterns, allowing ESP learners to receive targeted feedback on the language skills most relevant to their professional needs (Khampusaen et al., 2023). This is particularly important in the Thai context, where opportunities for authentic English interaction in professional settings may be limited.

5.3. Theoretical Contributions

This study contributes to theoretical understandings of L2 pronunciation acquisition in several ways. Beyond confirming SAT's principles in technology-enhanced environments, the findings provide empirical support for SLM-r's predictions regarding cross-linguistic phonetic similarity and new category formation. The differential effectiveness patterns observed across fricatives align closely with SLM-r's theoretical framework, demonstrating that technological intervention can systematically facilitate phonetic category establishment for sounds that present specific cross-linguistic challenges.

The research findings also demonstrate that the exponential learning curve proposed by SAT, where gains are most rapid during initial practice stages and gradually diminish (R. DeKeyser, 2007b), applies to technology-mediated learning environments (Taie, 2014; Kormos, 2006). The substantial improvements observed over the 16-week intervention period suggest that AI applications can accelerate early-stage learning through personalized feedback, particularly for the most challenging fricatives identified in this study. The findings also inform how individual differences in skill acquisition might be addressed through technology. SAT acknowledges that factors such as working memory capacity and phonological awareness influence skill development (Pawlak, 2022; Kormos, 2006; Skehan, 2015). The AI application's adaptive nature potentially accommodated these individual differences more effectively than traditional group instruction, allowing learners to progress at their own pace (Pawlak, 2022).

The segment-specific results provide evidence that technological intervention can optimize the noticing process described by R. Schmidt (1990) for different types of cross-linguistic challenges. For fricatives that showed significant improvement under AI-mediated instruction, the immediate feedback mechanisms appear to have enhanced learners' ability to notice discrepancies between their productions and target forms, particularly for sounds like /θ/, /ð/, /z/, and /ʃ/ that present specific perceptual or articulatory difficulties for Thai learners. This targeted enhancement of the noticing process may explain why certain fricatives benefited more from AI-mediated instruction than others, suggesting that the effectiveness of technological intervention varies systematically based on the underlying phonological constraints predicted by SLM-r and PAM-L2.

5.4. Limitations and Directions for Future Research

Despite its contributions, this study has several limitations that should be acknowledged. First, the 16-week intervention period, while substantial, may not have been sufficient to capture the long-term effects of AI-mediated pronunciation training. Future research should employ longitudinal designs to investigate whether the improvements observed in this study are maintained over time and transfer to spontaneous speech in authentic ESP contexts.

An important methodological limitation concerns the nature of the pronunciation test items employed in this study. The fricative sounds assessed in both pre-test and post-test evaluations were systematically practiced during the intervention period, meaning that participants received explicit instruction and repeated practice opportunities with the specific words and sentence contexts that subsequently appeared in the assessment instruments. While this approach ensured alignment between instruction and assessment and allowed for measurement of targeted learning outcomes, it necessarily limits the generalizability of the observed pronunciation improvements to untrained lexical items and novel communicative contexts.

This limitation can be understood through the lens of Skill Acquisition Theory, which distinguishes between context-dependent procedural knowledge and fully autonomous skill performance. Although participants demonstrated significant improvements in producing the trained fricative sounds within familiar lexical and sentential contexts, the extent to which these gains represent transferable pronunciation competence remains unclear. According to SAT's specificity principle, skills developed through practice in particular contexts may not automatically transfer to novel situations without additional training that promotes generalization across varied linguistic environments. Future studies should therefore incorporate transfer tasks that assess learners' ability to accurately produce target fricatives in untrained words, spontaneous discourse, and authentic communicative situations. Such investigations would better illuminate whether AI-mediated pronunciation

instruction facilitates the development of flexible, context-independent articulatory skills rather than memorization of specific lexical items.

While the study focused specifically on fricative sounds due to their importance and difficulty for Thai learners, pronunciation encompasses a broader range of segmental and suprasegmental features. Future studies should examine the effectiveness of AI-mediated instruction for other phonological aspects, such as vowel quality, consonant clusters, stress, rhythm, and intonation, which also contribute significantly to overall intelligibility.

6. Conclusions

This study investigated the effects of AI-mediated pronunciation training on Thai ESP learners' production of English fricative consonants. The findings indicate that while both AI-mediated and traditional instruction resulted in significant pronunciation improvements, the AI-based approach demonstrated superior effectiveness, particularly for fricative sounds that present the greatest cross-linguistic challenges for Thai learners (/θ/, /ð/, /z/, /ʃ/, and /h/). The results support the integration of AI technology into ESP pronunciation pedagogy, especially for addressing specific phonological difficulties related to learners' L1 background. The personalized, immediate feedback provided by AI applications appears to facilitate more efficient acquisition of challenging sounds compared to traditional instruction alone.

These findings contribute to our understanding of technology-enhanced pronunciation pedagogy and have important implications for ESP curriculum design and teaching practices in Thailand and similar EFL contexts. By leveraging AI technology to address specific pronunciation challenges, ESP programs can better prepare learners for effective communication in their professional domains. Future research should build on these findings by investigating the long-term effects of AI-mediated pronunciation training, its applicability to other phonological features, and its impact on communicative effectiveness in authentic professional contexts. As AI technology continues to evolve, its potential to enhance L2 pronunciation instruction, particularly for underserved populations and specific linguistic challenges, represents a promising frontier in applied linguistics and language pedagogy.

Author Contributions: Conceptualization, J.L.-u. and D.K.; methodology, J.L.-u. and D.K.; software, J.L.-u.; validation, J.L.-u.; formal analysis, J.L.-u.; investigation, J.L.-u.; data curation, J.L.-u.; writing—original draft preparation, J.L.-u. and D.K.; writing—review and editing, D.K.; visualization, J.L.-u.; supervision, D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Khon Kaen University (protocol code HE673271, date of approval 26 August 2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy and ethical considerations.

Acknowledgments: The authors would like to express their sincere gratitude to all participants who voluntarily devoted their time to this research. We also acknowledge the valuable feedback provided by the reviewers, whose constructive comments significantly strengthened this study.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. English Pronunciation Test Items

Section A: Individual Word Reading Task

Fricative /f/ (6 items)

1. Folk (initial position)
2. Performance (medial position)
3. Clef (final position)

Fricative /v/ (5 items)

4. Voice (initial position)
5. Movement (medial position)
6. Preserve (final position)

Fricative /θ/ (5 items)

7. Theatre (initial position)
8. Method (medial position)
9. Cloth (final position)

Fricative /ð/ (3 items)

10. They (initial position)
11. Rhythm (medial position)
12. Breathe (final position)

Fricative /s/ (6 items)

13. String (initial position)
14. Instrument (medial position)
15. Dance [final position]

Fricative /z/ (5 items)

16. Zither (initial position)
17. Musician (medial position)
18. Dulcimers (final position)

Fricative /ʃ/ (6 items)

19. Choreography (medial position)
20. Percussion (medial position)
21. Finish (final position)

Fricative /ʒ/ (4 items)

22. Culture (medial position)
23. Rouge (final position)
24. Homage (final position)

Fricative /h/ (4 items)

25. Rehearsal (initial position)

References

- Akhila, K., Nair, A. P., Jyothsana, K. N., Sreepriya, S., & Akhila, E. (2024). Fluttering into fluent English: Building an interactive voice-based AI learning app for language acquisition. *Journal of Artificial Intelligence and Capsule Networks*, 6(2), 158–170. [[CrossRef](#)]
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369–406. [[CrossRef](#)]
- Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology*, 105, 165–180. [[CrossRef](#)] [[PubMed](#)]
- Aryanti, R. D., & Santosa, M. H. (2024). A systematic review on artificial intelligence applications for enhancing EFL students' pronunciation skill. *The Art of Teaching English as a Foreign Language*, 5(1), 102–113. [[CrossRef](#)]

- Bashori, M., van Hout, R., Strik, H., & Cucchiari, C. (2021). Effects of ASR-based websites on EFL learners' vocabulary, speaking anxiety, and language enjoyment. *System*, 99, 102496. [\[CrossRef\]](#)
- Bashori, M., Van Hout, R., Strik, H., & Cucchiari, C. (2024). I can speak: Improving English pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching*, 18(5), 443–461. [\[CrossRef\]](#)
- Basturkmen, H. (2010). *Developing courses in English for specific purposes*. Palgrave Macmillan.
- Behr, N. S. (2022). English diphthong characteristics produced by Thai EFL learners: Individual practice using PRAAT. *Computer-Assisted Language Learning Electronic Journal*, 23, 401–424.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro, & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). John Benjamins.
- Cedar, P., & Termjai, M. (2021). Teachers' training of English pronunciation skills through social media. *Journal of Education and Innovation*, 23(3), 32–47.
- Chanaroke, U., & Niemprapan, L. (2020). The current issues of teaching English in Thai context. *Eau Heritage Journal Social Science and Humanities*, 10(2), 34–45.
- DeKeyser, R. (2007a). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.
- DeKeyser, R. (2007b). Skill acquisition theory. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition* (pp. 97–113). Routledge.
- DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, 19(2), 195–221. [\[CrossRef\]](#)
- DeKeyser, R. M. (2009). Cognitive-psychological processes in second language learning. In M. Long, & C. Doughty (Eds.), *Handbook of Second Language Teaching* (pp. 119–138). Wiley-Blackwell.
- DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 94–112). Routledge.
- Dennis, N. K. (2024). Using AI-powered speech recognition technology to improve English pronunciation and speaking skills. *IAFOR Journal of Education: Technology in Education*, 12(2), 107–126. [\[CrossRef\]](#)
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.
- Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27(2), 305–352. [\[CrossRef\]](#)
- Flege, J. E. (1995). Second-language speech learning: Theory, finding, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issue in cross-language research* (pp. 229–273). York Press.
- Flege, J. E., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge Press. [\[CrossRef\]](#)
- Guskaroska, A. (2019). *ASR as a tool for providing feedback for vowel pronunciation practice* [Master's thesis, Iowa State University]. [\[CrossRef\]](#)
- Huang, Q., & Kitikanan, P. (2022). Production of the English "sh" by L2 Thai learners: An acoustic study. *Theory and Practice in Language Studies*, 12(8), 1508–1515. [\[CrossRef\]](#)
- Isarankura, S. (2016). Using the audio-articulation method to improve EFL learners' pronunciation of the English /v/ sound. *Thammasat Review*, 18(2), 116–137.
- Jenkins, J. (2000). *The phonology of English as an international language: New models, new norms, new goals*. Oxford University Press.
- Kachinske, I. (2021). Skill acquisition theory and the role of rule and example learning. *Journal of Contemporary Philology*, 4(2), 25–41. [\[CrossRef\]](#)
- Kaewpet, C. (2009). A framework for investigating learner needs: Needs analysis extended to curriculum development. *Electronic Journal of Foreign Language Teaching*, 6(3), 209–220.
- Kang, S. W., Lee, Y. J., Lim, H. J., & Choi, W. K. (2024). Development of AI convergence education model based on machine learning for data literacy. *Advanced Industrial Science*, 3(1), 1–16.
- Kanokpermpoon, M. (2007). Thai and English consonantal sounds: A problem or a potential for EFL learning? *ABAC Journal*, 27(1), 57–66.
- Kaya, S. (2021). From needs analysis to development of a vocational English language curriculum: A practical guide for practitioners. *Journal of Pedagogical Research*, 5(1), 154–171. [\[CrossRef\]](#)
- Khamkhen, A. (2010). Teaching English speaking and English speaking tests in the Thai context: A reflection from Thai perspective. *English Language Teaching*, 3(1), 184. [\[CrossRef\]](#)
- Khampusaen, D., Chanprasopchai, T., & Lao-un, J. (2023). Empowering Thai community-based tourism operators: Enhancing English pronunciation abilities with AI-based lessons. *Mekong Journal*, 21(1), 45–60.

- Kitikanan, P. (2017). The effects of L2 experience and vowel context on the perceptual assimilation of English fricatives by L2 Thai learners. *English Language Teaching*, 10(12), 72. [\[CrossRef\]](#)
- Kormos, J. (2006). *Speech production and second language acquisition*. Lawrence Erlbaum Associates.
- Lao-un, J., & Bunyaphithak, W. (2025). Needs analysis of English language in Thai performing arts profession. *Asia Social Issues*, 18(5), 2–13.
- Lara, M. S., Subhashini, R., Shiny, C., Lawrance, J. C., Prema, S., & Muthuperumal, S. (2024, April 12–14). *Constructing an AI-assisted pronunciation correction tool using speech recognition and phonetic analysis for ELL*. 2024 10th International Conference on Communication and Signal Processing (ICCSP) (pp. 1021–1026), Melmaruvathur, India. [\[CrossRef\]](#)
- Lee, C., Zhang, Y., & Glass, J. (2013, October 18–21). *Joint learning of phonetic units and word pronunciations for ASR*. The 2013 Conference on Empirical Methods in Natural Language Processing (pp. 182–192), Seattle, WA, USA. [\[CrossRef\]](#)
- Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press.
- Liu, Y., binti Ab Rahman, F., & binti Mohamad Zain, F. (2025). A systematic literature review of research on automatic speech recognition in EFL pronunciation. *Cogent Education*, 12(1), 2466288. [\[CrossRef\]](#)
- McCrocklin, S. (2019). Dictation programs for second language pronunciation learning. *Journal of Second Language Pronunciation*, 5(2), 242–258. [\[CrossRef\]](#)
- McCrocklin, S., Fetting, C., & Markus, S. (2022, September 14). *SalukiSpeech: Integrating a new ASR tool into students' English pronunciation practice*. Virtual PSLT, Virtual. [\[CrossRef\]](#)
- Mohammadkarimi, E. (2024). Exploring the use of artificial intelligence in promoting English language pronunciation skills. *LLT Journal: A Journal on Language and Language Teaching*, 27(1), 98–115. [\[CrossRef\]](#)
- Naruemon, D., Bhoomanee, C., & Pothisuwan, P. (2024). Challenges of Thai learners in pronouncing English consonant sounds. *The Golden Teak: Humanity and Social Science Journal*, 30(3), 1–19.
- Neri, A., Cucchiarini, C., & Strik, W. (2003, August 3–9). *Automatic Speech Recognition for second language learning: How and why it actually works*. 15th ICPhS (pp. 1157–1160), Barcelona, Spain.
- Ngo, T. T. N., Chen, H. H. J., & Lai, K. K. W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 36(1), 4–21.
- Noom-ura, S. (2013). English-teaching problems in Thailand and Thai teachers' professional development needs. *English Language Teaching*, 6(11), 139. [\[CrossRef\]](#)
- Pawlak, M. (2022). Research into individual differences in SLA and CALL: Looking for intersections. *Language Teaching Research Quarterly*, 31, 200–233. [\[CrossRef\]](#)
- Peerachachayane, S. (2022). Towards the phonology of Thai English. *Academic Journal of Humanities and Social Sciences Burapha University*, 30(3), 64–92.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. [\[CrossRef\]](#)
- Rajkumari, Y., Jegu, A., Fatma, D. G., Mythili, M., Vuyyuru, V. A., & Balakumar, A. (2024, October 25–26). *Exploring neural network models for pronunciation improvement in English language teaching: A pedagogical perspective*. 2024 International Conference on Intelligent Systems and Advanced Applications (ICISAA) (pp. 1–6), Pune, India. [\[CrossRef\]](#)
- Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, 63(3), 595–626. [\[CrossRef\]](#)
- Schmidt, R. (1990). The noticing hypothesis: Awareness and second language acquisition. *Studies in Second Language Acquisition*, 12(3), 1–20.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press. [\[CrossRef\]](#)
- Schmidt, T., & Strassner, T. (2022). Artificial intelligence in foreign language learning and teaching. *Anglistik*, 33(1), 165–184. [\[CrossRef\]](#)
- Segalowitz, N. (2010). *Cognitive bases of second language fluency* (1st ed.). Routledge. [\[CrossRef\]](#)
- Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford University Press.
- Shafiee Rad, H., & Roohani, A. (2024). Fostering L2 learners' pronunciation and motivation via affordances of artificial intelligence. *Computers in the Schools*, 42(3), 1–12. [\[CrossRef\]](#)
- Skehan, P. (2015). Foreign language aptitude and its relationship with grammar: A critical overview. *Applied Linguistics*, 36(3), 367–384. [\[CrossRef\]](#)
- Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39(4), 456–466. [\[CrossRef\]](#)
- Sukying, A., Supunya, N., & Phusawisot, P. (2023). ESP teachers: Insights, challenges and needs in the EFL context. *Theory and Practice in Language Studies*, 13(2), 396–406. [\[CrossRef\]](#)
- Taie, M. (2014). Skill acquisition theory and its important concepts in SLA. *Theory and Practice in Language Studies*, 4(9), 1971–1976. [\[CrossRef\]](#)

- Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, 36(3), 326–344. [[CrossRef](#)]
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Wang, J. (2024). Optimizing English pronunciation teaching through motion analysis and intelligent speech feedback systems. *Molecular & Cellular Biomechanics*, 21(4), 652. [[CrossRef](#)]
- Wongsuriya, P. (2020). Improving the Thai students' ability in English pronunciation through mobile application. *Educational Research and Reviews*, 15(4), 175–185.
- Wulff, S., & Ellis, N. C. (2018). Usage-based approaches to second language acquisition. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: The state of the science across its subfields* (pp. 37–56). John Benjamins Publishing Company. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.