PASAA Journal
Volume 69, July – December 2024, 370–412

PASAA

Exploring Gender Differences in Correcting Mother Tongue Influence in EFL Using CALL: A Study with Thai Undergraduate Students

Simon Moxon
Walailak University, Nakhon Si Thammarat, Thailand
Corresponding author: simon@sitamoxon.com

Article information	
Abstract	This study examined the pronunciation challenges faced by Thai EFL learners due to phonetic and phonological differences between Thai and English. It evaluates "ALL-Talk," a custom-built Automatic Speech Recognition (ASR) tool that provides individualized feedback through example speech, waveforms, and corrective cues. Additionally, the study investigated gender-based differences in pronunciation improvement, responding to literature that suggests varying impacts of gender on language acquisition. Using a quasi-experimental design over a 10-week intervention period, the research involved Thai undergraduate EFL students (n = 96, female = 73, male = 23, mean age = 18.4 years). Participants practiced autonomously with ALL-Talk, and pre- and post-assessments measured improvements in problematic phonemes identified in the literature and baseline tests. The data were analyzed using ANCOVA to control for initial proficiency differences, with paired samples t-tests applied to evaluate phoneme-specific improvements. A regression analysis showed that specific features of ALL-Talk (e.g., Login, Text-To-Speech, Evaluation, and Task Review) did not individually predict pronunciation improvement, though overall usage positively impacted performance. Results indicated that, although students improved overall, they encountered persistent difficulty with the phonemes /ʊ/ and

E-ISSN: 2287-0024

	/ʧ/, even with similar sounds present in Thai, indicating a more complex influence of phonetic context. The findings highlight the potential of ASR technology for personalized pronunciation training in EFL learning and suggest that gender-specific strategies may further enhance effectiveness, as female students showed greater engagement with visual feedback features. Future research should examine phonetic context effects on pronunciation and ASR's role in providing targeted, autonomous feedback across different learner demographics.
Keywords	EFL pronunciation, computer-aided language learning (CALL), mother tongue phonetic interference, autonomous language learning, gender differences in language acquisition
APA citation:	Moxon, S. (2024). Exploring gender differences in correcting mother tongue influence in EFL using CALL: A study with Thai undergraduate students. *PASAA, 69*, 369–412.

1. Introduction

Learning English as a Foreign Language (EFL) poses distinct challenges, particularly in pronunciation, mainly due to the influence of the learner's native language (L1). This influence, commonly referred to as mother tongue influence (MTI), stems from disparities in the phonetic and phonological structures between the native language and English (Gabriel, 2023; Jahandar et al., 2012; Ngo et al., 2023). For Thai EFL learners, specific challenges include difficulties with English sounds absent in Thai, such as certain fricatives and final consonant clusters, which often lead to substitutions, omissions, or vowel insertions to align with Thai phonotactic rules (Behr, 2022; Kapranov, 2022; Sridhanyarat, 2017). Moreover, early-stage EFL learners often transfer their native phonetic habits into English (Behr, 2022; Jahandar et al., 2012; Lai et al., 2009; van den Doel et al., 2018).

This study explored the remedial potential of technology, particularly Automatic Speech Recognition (ASR), in addressing these pronunciation challenges. Despite MTI's impact on pronunciation accuracy, pronunciation training is often marginalized in traditional EFL curricula, partly due to limited class time and lack of teacher confidence in teaching pronunciation explicitly (Thomson & Derwing, 2014). Recent advancements in technology, particularly ASR, offer promising solutions for overcoming MTI by providing learners with autonomous, consistent, and targeted feedback (Behr, 2022; Haggag, 2018). ASR tools can analyze student speech in real time, highlighting specific pronunciation errors and offering corrective feedback without the potential anxiety of peer comparison (Huang & Jia, 2016; Moxon, 2021).

This study contributes to the literature by introducing "ALL-Talk" (Moxon, 2024), an ASR tool specifically designed for Thai EFL learners. ALL-Talk advances ASR applications by embedding pronunciation practice within authentic language contexts, moving beyond isolated phoneme drills. It integrates example sentences and conversational phrases, allowing learners to practice sounds as part of complete linguistic units, which promotes both accuracy and fluency. Additionally, ALL-Talk offers visual waveform feedback across entire sentences and targeted corrective cues within realistic language settings, supporting learners in making meaningful adjustments and self-correcting in real-time. Through these features, ALL-Talk provides a more holistic pronunciation practice experience, bridging the gap between technical accuracy and practical language use for Thai EFL learners.

Additionally, this study examined the role of gender in pronunciation improvement, a relatively underexplored area in ASR-based pronunciation practice. Prior research suggests gender may influence pronunciation learning trajectories due to differences in learning strategies and engagement with feedback (Bryla-Cruz, 2021; Jahandar et al., 2012). By investigating gender-specific interactions with ASR feedback, this study responds to Moxon's (2021) call for more research into gender dynamics in autonomous pronunciation practice, offering insights that could shape more targeted ASR applications for diverse learner profiles.

The following research questions guided this study:

1. Which English sounds represent the greatest challenges for Thai EFL students?

2. To what extent does autonomous learning through an ASR platform reduce the influence of mother tongue pronunciation?

3. How does gender influence the effectiveness of ASR in counteracting mother tongue pronunciation influences?

2. Literature Review

2.1 Language Differences

The Thai phonetic inventory poses distinct challenges for Thai EFL learners. Thai has 44 consonants and 32 vowels (Becker, 1995). These characters produce 21 distinct consonant sounds (Kanokpermpoon, 2007) and nine pure vowel sounds (Becker, 1995). Thai has only three fricative sounds, which are acoustically similar to the English fricatives: /s/, /h/, and /f/ (Isarankura, 2015; Tingsabadh & Abramson, 1993). According to Isarankura (2015), the three fricatives occur only as initial syllable sounds. Among the eight final consonant sounds, three (/m/, /n/, /ŋ/) are nasals, two (/w/, /j/) are approximants, and three (/p/, /t/, /k/) are plosives. A lack of audible plosive release means there are no audible final fricative sounds (Becker, 1995; Isarankura, 2015).

In contrast, English has 21 consonants and five vowels, typically producing 24 distinct consonant sounds and 20 vowel sounds depending on dialect (Mantei et al., 2021). This dialect variation contributes to the difficulties faced by EFL learners.

English consonant clusters also pose a significant problem for EFL learners. The English language offers up to three consonants in initial consonant cluster sounds, which typically begin with /s/ or /t/, such as “street” and “through,”, and up to four in final consonant cluster sounds, such as “twelfths.” In contrast, the Thai language has few initial consonant cluster sounds, comprising of no more than two consonants. While there are similar consonant cluster sounds in the two languages, such as /kr/, /kl/, and /kw/, the Thai language has no equivalent initial consonant cluster sounds for voiced and voiceless /th/ (/ð/ and / θ /, respectively) (Jahandar et al., 2012).

2.2 Mother Tongue Influences

L1 interference in L2 pronunciation is well-documented (e.g., Fauzi, 2021; Fitriani & Zulkarnain, 2019; Qader et al., 2023; Sridhanyarat, 2017), especially when English phonemes are absent in the learners' native language (Behr, 2022; Gabriel, 2023; Isarankura, 2015; Jahandar et al., 2012; van den Doel et al., 2018), or the consonant clusters do not align with Thai phonotactic rules (Lai et al., 2009). However, previous research has demonstrated various degrees of association between challenges in L2 pronunciation and the phonetic characteristics of the L1.

Studying 16 CEFR B2 level EFL Norwegian students, Kapranov (2022) deduced that L2 learners instinctively align the phonetic elements they encounter in L2 with the phonetic categories of the L1. This study, which concentrated on the participants' proficiency in distinguishing between the /s/ and /z/ sounds, required each student to transcribe two brief movie summaries using the International Phonetic Alphabet (IPA). The results found that over 75% of the students replaced /z/ with /s/.

Foundational theories like Lado's (1957) contrastive analysis hypothesis (CAH) and Eckman's (1977) markedness differential hypothesis (MDH) provide a theoretical backdrop for understanding these L1-L2 phonetic discrepancies. These frameworks suggest that L2 phonemes that are marked or absent in the L1 are harder for learners to acquire. This premise is supported by Flege's (1995) speech learning model (SLM), which theorizes that non-native speakers perceive phonetic sounds differently from native speakers.

In a study by Sridhanyarat (2017), the MDH was employed to investigate Thai undergraduates' acquisition and substitution patterns of English marked and unmarked fricatives. Participants (N = 45) were categorized into three proficiency levels based on their national test scores and presented with 12 oral interview questions, followed by two reading tasks. The findings revealed that, except for advanced-level participants, English fricative sounds /f/, /s/, /θ/, /ð/, /z/, /v/, /ʃ/, and /ʒ/ were often replaced with simpler, unmarked sounds. The study did not encompass affricates, vowels, medial fricatives, or gender-related phoneme acquisition differences.

The influences of MDH were also observed by Behr (2022), who found that Thai students could proficiently articulate English diphthongs that could be derived from Thai monophthongs but struggled with English sounds absent in Thai. Relating to the SLM, Jahandar et al. (2012) and Lai et al. (2009) noted that EFL learners often inserted vowel sounds into consonant clusters incompatible with their L1 phonetic system, a pattern prevalent among the students in Behr's study.

Recent advancements in technology, particularly ASR, offer promising solutions for overcoming MTI by providing learners with autonomous, consistent, and targeted feedback (Behr, 2022; Haggag, 2018). ASR tools can analyze student speech in real time, highlighting specific pronunciation errors and offering corrective feedback without the potential anxiety of peer comparison (Huang & Jia, 2016; Moxon, 2021).

2.3 Pedagogical Influences

Thomson and Derwing (2014) argue that pronunciation is often marginalized in language teaching. This neglect is partly due to the time dedicated to language instruction, student-teacher ratios, and a lack of confidence and expertise in effectively teaching pronunciation, a gap emphasized by Nushi and Sadeghi (2021). Recent pedagogical trends have shifted from traditional drill-based methods towards integrating pronunciation into communicative practice (Afshari & Ketabi, 2017; Alomari, 2024; Sun, 2023).

2.4 Corrective Feedback

Studies have highlighted the importance of corrective feedback and how its method of delivery can greatly affect student confidence, anxiety, and motivation (Huang & Jia, 2016; Moxon, 2021; Wiboolyasarin et al., 2023; Zou et al., 2023). For instance, feedback given in front of peers can cause anxiety and loss of confidence to the recipient (Hinks, 2003; Huang and Jia, 2016; Wiboolyasarin et al., 2023). Moreover, the literature shows that feedback only benefits the student if its meaning can be interpreted. For instance, Moxon (2021) argues that an overall evaluation score is ineffective for students when identifying specific errors. He concludes that feedback should pinpoint the error and offer corrective instruction.

Regarding corrective feedback, Behr (2022) and Olson and Offerman (2021) advocate the visual feedback paradigm, which they claim leads to significant pronunciation improvement. Behr (2022) used spectrograms generated by PRAAT for Thai students to analyze their pronunciation of eight English diphthongs. The study found that almost 77% of participants claimed visual feedback assisted them in identifying the two-vowel movement of the diphthongs. Olson and Offerman (2021) compared three studies that used visual feedback when practicing L2 segmental pronunciation. They found that both forms of visual feedback (waveforms and spectrograms) effectively improved L2 pronunciation.

2.5 ASR Technology

Technology integration in educational contexts has opened pathways for mitigating L1 influences on L2 pronunciation. ASR has gained recognition for its efficacy in improving pronunciation among EFL learners. Research has highlighted its significant role in fostering students' engagement (Behr, 2022; Haggag, 2018) and refining the pronunciation precision of learners from diverse linguistic backgrounds (Kayaoğlu, 2019; Moxon, 2021; Ngo et al., 2023; Sun, 2023).

While literature offers growing support for ASR integration into educational contexts, there remains debate on its efficacy, suitability, and reliability. A meta-analysis by Ngo et al. (2023) found that ASR had a medium effect size on pronunciation performance, leading to significant improvements on segmental pronunciation, but smaller effects on suprasegmental pronunciation. The authors suggest that ASR is more effective in EFL contexts for adult learners or those at an intermediate proficiency level.

Spring and Tabuchi (2022) echo similar findings regarding the consistency of ASR's effectiveness for pronunciation. Their study, which examined the impact of guided practice and treatment duration, concluded that while treatment length did not significantly influence outcomes, ASR's effects on pronunciation varied markedly among individual learners.

Moxon (2024) identifies a limitation of ASR, arguing that exclusive reliance on the technology can lead to inaccurate results and misleading feedback, as ASR may mistakenly interpret a mispronounced word as a similar-sounding one, such as "fine" instead of "find." ALL-Talk seeks to address this issue by comparing the submitted speech directly with the target text, reducing the likelihood of such errors.

Regarding ASR's suitability as an autonomous EFL learning tool, Moxon (2021) argues that its effectiveness depends on the technology's ability to deliver meaningful feedback that learners can interpret and apply independently. In his review of Pronunciation Coach 3D, Moxon (2023) emphasizes that numerical feedback alone, particularly at phrase or word level, is insufficient for students to accurately identify and correct phonetic level errors. The interpretability of feedback and technical expertise required for other ASR platforms, such as PRAAT, remain ongoing points of debate in the literature (e.g., Brett, 2004; Olson, 2014; Setter & Jenkins, 2005).

Despite the advancements in ASR technology, there remains a notable gap in research concerning the effectiveness of ASR tools that provide feedback at multiple linguistic units (phrase, word, syllable, and phoneme) to support autonomous pronunciation improvement. The present study addressed this gap by examining the potential of ALL-Talk, which integrates a comparative feedback mechanism at these linguistic units, aiming to enhance pronunciation accuracy among EFL learners and offer a more comprehensive solution to the limitations observed in existing ASR-based learning tools.

2.6 Gender Differences

According to the literature, gender differences in the context of EFL pronunciation are indeterminate. Jahandar et al. (2012) and Hariri (2012) argue that gender does not significantly impact L2 pronunciation, although females generally perform better than males. In contrast, Khamkhien (2010) found that females outperformed males in word stress identification, suggesting that gender differences may exist within specific linguistic tasks. Moreover, Bryla-Cruz (2021) argues that social influences rather than biological diversities drive gender differences in L2 learning.

There remains a notable gap in the literature regarding how gender differences manifest in the context of using technology for pronunciation training. This oversight is especially pertinent in studies involving Thai EFL learners, where cultural and educational dynamics can shape gender-specific approaches to language learning. This gap underscores the need for more targeted research to understand the interplay of gender and technology in L2 learning.

2.7 Conclusion

The literature underscores the complex interplay of linguistic, pedagogical, and technological factors in EFL pronunciation practice. This research aimed to contribute valuable insights into the efficacy of pronunciation applications for autonomous language learning by situating this study within these ongoing debates and leveraging current technological advancements.

3. Methodology

3.1 Research Design

The present study employed a quasi-experimental design over a 10-week intervention period to evaluate the efficacy of ALL-Talk in enhancing pronunciation accuracy. Pronunciation improvements were assessed by comparing pre-test and post-test results, focusing specifically on phonemes identified as problematic in both the literature and the pre-test assessments. Additionally, this study examined the influence of gender on pronunciation improvement, exploring potential differential effects.

3.2 Participants

The participants were Thai EFL undergraduate students from a university in southern Thailand (n = 96, female = 73, male = 23, mean age = 18.4 years). The greater proportion of female participants was due to the larger number of female students enrolled in the program.

3.3 Instruments

The primary research instrument was ALL-Talk, a web-based ASR application specifically developed to support autonomous pronunciation practice. ALL-Talk provides several interactive features, including example speech, visual waveform feedback, and animated corrective instructions. The tool quantifies pronunciation accuracy at multiple linguistic levels—phrase, word, syllable, and phoneme—and offers color-coded feedback to indicate accuracy levels. Details of ALL-Talk's main functions are included in Appendix B.

To assess pronunciation improvements, pre-test and post-test scores were collected through the application at the beginning and end of the 10-week intervention. Each test comprised 30 short phrases designed to encompass all 44 English phonemes. Unbeknownst to the participants, these phrases were kept consistent across tests, with phrase order randomized to minimize memorization or practice effects. The assessment design, adapted from Mahzoun and Han (2019), specifically targeted phonemic challenges such as initial, middle, and final English fricatives and consonant clusters —areas often mispronounced due to L1 phonotactic constraints, as documented in the literature (e.g., Behr, 2022; Sridhanyarat, 2017).

Pre- and post-test assessments, alongside intervention practice tasks, were completed autonomously by participants through ALL-Talk at their convenience. For each assessment phrase, participants were presented with the target phrase visually and prompted to record their pronunciation. They also had the option to listen to each phrase via ALL-Talk's Text-To-Speech (TTS) feature, which provides authentic nativelike speech in a range of male and female speech characters. Both the TTS audio and participants' recordings were automatically converted to waveforms, enabling a visual comparison of target and recorded speech (see Appendix B, Figure 1). Participants could adjust the articulation rate and pitch of the target speech to suit their listening preferences. Additionally, they could review and reattempt each phrase before submitting it for evaluation. Upon completing all test phrases, participants were directed to the main page, where they could view assessment scores and review detailed analysis results for each phrase (see Appendix B, Figures 2 and 3).

Intervention tasks, while distinct from the pre- and post-test assessments, were based on the same problematic phonemes, consonant clusters, and the use of initial, medial, and final sounds as the pre- and post-test assessments, following a similar methodology in selecting and arranging targeted phonemic challenges. However, intervention tasks could be repeated at participants' discretion, whereas pre- and post-test assessments were limited to a single attempt. For each phoneme score, participants could click on the relevant phoneme to access written and animated guidance on correct articulation (see Appendix B, Figure 4).

3.4 Data Collection

Data collection occurred throughout the 2023-2024 academic year. Participants were selected via convenience sampling and received detailed information about the study's objectives and their right to withdraw at any stage. Between pre-test and post-test, they were assigned ten weekly speaking tasks, each consisting of ten short phrases. Additionally, they could enter and practice their own words and phrases at will. Evaluation scores and interaction with the main features of ALL-Talk were automatically captured by the application and stored on a secured server.

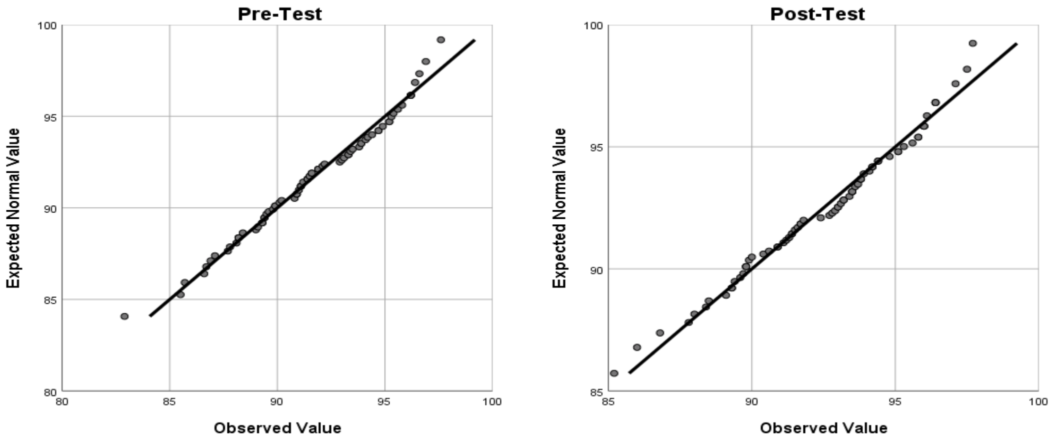
3.5 Data Preparation

Data were first screened to ensure students had fully completed pre-test and post-test assessments. Eighteen instances of incomplete data were identified and removed. Descriptive statistics were then explored to identify any erroneous or outlier data. Five outlier cases were detected where recording issues rather than genuine pronunciation problems were evident. These were deemed illegitimate scores and excluded from the analysis (Osborne & Overbay, 2004).

A Shapiro-Wilk test showed pre-test and post -test scores did not significantly depart from normality: $W(73) = .983$, $p = .414$ and $W(73) = .980$, $p = .294$, respectively (see Figure 1). Skewness of pre-test data ($X(73) = -.24$, $z = -.84$) and post-test data ($X(73) = -.33$, $z = -1.16$) as well as Kurtoses of pre-test data ($X(73) = -.40$, $z = -.72$) and post-test data ($X(73) = -.40$, $z = -.72$) were all found to be within acceptable limits (Ghasemi & Zahediasl, 2012), suggesting the assumptions of parametric testing had been met.

Figure 1

Q-Q Plots of Pre-Test and Post-Test Scores



Descriptive statistics were used to categorize students into quartile-range groups according to their initial proficiency levels, as indicated by their pre-test scores. These groups were subsequently used for additional analysis (Sridhanyarat, 2017).

3.6 Justification for Statistical Techniques

3.6.1 Descriptive Statistics and Phoneme Scores Below 75%

Descriptive statistics were used to identify phonemes with pre-test scores below 75%, establishing a baseline to assess which phonemes were most problematic for the participants. This threshold guided the identification of targeted phonemic challenges and informed subsequent analyses of improvement. This foundational approach aids in understanding the general patterns within the data and is commonly recommended as an initial step for exploring and summarizing datasets (Field, 2013).

3.6.2 ANCOVA for Overall Improvement

ANCOVA was applied to evaluate overall improvement in pronunciation while controlling for baseline proficiency. This method allows for the adjustment of initial proficiency levels, providing a clearer view of the impact of the intervention on pronunciation gains. By controlling for pre-test scores, ANCOVA isolates the effect of the intervention, ensuring that improvements are attributed to ALL-Talk rather than baseline differences among participants. ANCOVA is widely used in educational research for this purpose (Tabachnick & Fidell, 2013).

3.6.3 Paired Samples *t*-tests for Pre-post Improvements

Paired samples *t*-tests assessed overall improvements from pre-test to post-test for the entire sample and by proficiency level. This test was chosen for its effectiveness in measuring significant mean differences in paired observations, enabling a focused analysis of changes due to the intervention across different skill levels. Paired samples t-tests are robust for measuring significant mean differences in paired observations, making them well-suited for pre- and post-test comparisons within the same group (Cohen, 1988).

3.6.4 ANOVA with Bonferroni Post Hoc

To compare improvement across proficiency levels, ANOVA with Bonferroni post hoc adjustments was applied. ANOVA is effective for detecting statistically significant differences across groups, while the Bonferroni adjustment helps control for Type I error in multiple comparisons, providing a more conservative approach (Field, 2013).

3.6.5 Paired Samples *t*-tests with Holm-Bonferroni Correction for Phoneme-specific Changes

Separate paired samples *t*-tests with Holm-Bonferroni adjustments were used to compare pre-test and post-test scores for each phoneme, targeting specific improvements. The Holm-Bonferroni correction controls for the false discovery rate, which is crucial when conducting multiple comparisons to prevent inflation of Type I errors (Holm, 1979).

3.6.6 Multiple Linear Regression for System Usage Impact

Finally, multiple linear regression was employed to examine the effect of four areas of ALL-Talk system usage (e.g., TTS engagement, waveform analysis) on overall pronunciation improvement. This approach is suitable for analyzing relationships between continuous variables, allowing the study to assess which features of ALL-Talk contributed most significantly to pronunciation gains (Pedhazur, 1997).

These statistical techniques collectively allow for a comprehensive evaluation of the intervention, from general improvements to phoneme-specific gains and feature-based effectiveness. Additionally, grouping students by quartiles based on pre-test scores facilitates detailed analysis by proficiency level, providing a more comprehensive understanding of how the intervention affects different learner groups (Sridhanyarat, 2017).

4. Results/Findings

4.1 Research question 1: Which English sounds represent the greatest challenges for Thai EFL students?

To investigate the phonemes that Thai EFL students find most challenging, descriptive analysis of pre-test scores was used to identify phonemes with average accuracy rates below 75%. Results were computed based on overall performance and proficiency level.

The descriptive analysis revealed seven phonemes with overall mean scores below 75% (/d/, /ʤ/, /g/, /ʧ/, /ʊ/, /z/, and /θ/). Of these, five (/g/, /ʤ/, /θ/, /z/, and /ʧ/) concurred with the known problematic sounds identified in the literature (Becker, 1995; Sridhanyarat, 2017), which include /s/, /f/, /ʃ/, /v/, /z/, /θ/, /ð/, /ʒ/, /g/, /ʤ/, /ʧ/, /r/, and /l/ (see Table 1).

Table 1

Pre-test Descriptive Information for Problematic Phonemes

Phoneme	*n*	Min	Max	*M*	*SD*
d	73	54.7	93.0	73.27	8.94
ð	73	87.9	99.3	95.06	2.47
ʤ	73	31.8	95.9	68.10	13.99
f	73	71.7	98.9	89.19	5.30
g	73	34.4	86.0	62.15	10.60
l	73	71.2	97.4	85.43	6.43
r	73	65.5	94.7	82.65	7.28
s	73	53.7	93.4	76.76	10.33
ʃ	73	77.6	100.0	93.20	5.00
ʧ	73	50.8	95.4	73.97	9.98
ʊ	73	40.8	76.6	53.63	7.50
v	73	68.5	99.9	90.77	6.03
z	73	49.8	95.1	73.16	12.22
ʒ	73	56.4	100.0	85.74	11.08
θ	73	44.3	88.1	69.54	8.63

As illustrated in Appendix A, Table 1, the results revealed two phonemes, /g/ and /ʊ/, with mean scores below 75% consistently across all proficiency levels, with /ʊ/ notably the most problematic.

4.2 Research question 2: To what extent does autonomous learning through an ASR platform reduce the influence of mother tongue pronunciation?

4.2.1 Analysis of overall pronunciation accuracy

An Analysis of Covariance (ANCOVA) was conducted to account for initial pronunciation performance and control for baseline differences in proficiency. The model included the mean improvement score as the dependent variable, gender as the fixed factor, and pre-test scores as the covariate. As Table 2 shows, the results revealed that controlling for starting proficiency, the model significantly predicted post-test scores ($F(2, 70) = 13.31$, $p < .001$), explaining 28% of the variance in post-test performance ($\eta^2 = .28$). The results suggest that initial pronunciation proficiency influenced post-test performance.

Table 2

ANCOVA Results for Pronunciation Improvement, Controlling for Pre-Test Scores

Source	*SS*	*df*	*MS*	*F*	*p*	Noncent Parameter	Observed Power [b]
Corrected Model	123.92 [a]	2	61.96	13.31	.000	26.62	.997
Intercept	129.31	1	129.31	27.78	.000	27.78	.999
Pre-Test Score	123.92	1	123.92	26.62	.000	26.62	.999
Gender	2.00	1	2.00	.41	.525	.41	.097
Error	327.74	71	4.62				
Total	503.27	73					
Corrected Total	449.76	72					

a. $\eta^2 = .28$

b. Computed using alpha = .05

4.2.2 Pronunciation improvement by proficiency level

Investigating improvements in pronunciation based on proficiency level, a series of paired samples t-tests were conducted (see Table 3). Overall pronunciation improvement was significant (p = .005). However, at proficiency level, overall improvement scores were significant only for lower and mean quartile range students. Conversely, students in the upper quartile group appear to have deteriorated between pre-test and post-test assessments.

Table 3

Paired Samples t-Test Comparing Pre-test and Post-test Pronunciation Scores Across Proficiency Levels and Genders

Proficiency Level	Gender	*MD*	*SD*	*SEM*	*t*	*df*	*P*	95% Confidence Interval	
								Lower Bound	Upper Bound
All	Both	.86	2.50	.29	2.93	72	.005*	.27	1.44
	Male	.86	2.02	.50	1.71	15	.108	-.21	1.94
	Female	.85	2.63	.35	2.45	56	.018*	.16	1.55
Lower Quartile	Both	2.46	2.79	.64	3.85	18	.001*	1.12	3.81
	Male	2.65	2.48	1.24	2.14	3	.122	-1.29	6.59
	Female	2.41	2.94	.76	3.17	14	.007*	.78	4.04
Mean	Both	.69	1.94	.33	2.11	34	.042*	.03	1.36
	Male	.23	1.90	.78	.30	5	.776	-1.77	2.23
	Female	.79	1.97	.37	2.15	28	.040*	.04	1.53
Upper Quartile	Both	-.45	2.35	.54	-.83	18	.418	-1.58	.69
	Male	.30	1.22	.50	.60	5	.574	-.98	1.58
	Female	-.79	2.70	.75	-1.06	12	.310	-2.42	.84

* $p < .05$, 2-tailed.

A one-way Analysis of Variance (ANOVA) was conducted to investigate the differences in improvement across proficiency levels (see Table 4). The results revealed a significant difference between groups: $F(2) = 7.84$, $p = .001$, $\eta^2 = 0.18$. Bonferroni Post Hoc tests indicated significant differences between the lower and mean quartile groups, and between the lower and upper quartile groups. No significant difference was found between the mean and upper quartile groups.

Table 4

Bonferroni Post Hoc Comparisons of Pronunciation Improvement Across Quartile Range Groups and Genders

Gender	(I) Proficiency Level	(J) Proficiency Level	*MD*	*SE*	*p*	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Both	Lower Quartile	Mean	1.77	.65	.025*	.17	3.37
		Upper Quartile	2.91	.74	.001*	1.09	4.73
	Mean	Lower Quartile	-1.77	.65	.025*	-3.37	-.17
		Upper Quartile	1.14	.65	.257	-.46	2.74
	Upper Quartile	Lower Quartile	-2.91	.74	.001*	-4.74	-1.09
		Mean	-1.14	.65	.257	-2.74	.46
Male	Lower Quartile	Mean	2.42	1.19	.188	-.84	5.68
		Upper Quartile	2.35	1.19	.208	-.91	5.61
	Mean	Lower Quartile	-2.42	1.19	.188	-5.68	.84
		Upper Quartile	-.07	1.06	1.000	-2.98	2.85
	Upper Quartile	Lower Quartile	-2.35	1.19	.208	-5.61	.91
		Mean	.07	1.06	1.000	-2.85	2.98
Female	Lower Quartile	Mean	1.63	0.77	0.118	-0.28	3.53
		Upper Quartile	3.21	0.92	0.003*	0.94	5.47
	Mean	Lower Quartile	-1.63	0.77	0.118	-3.53	0.28
		Upper Quartile	1.58	0.81	0.168	-0.42	3.58
	Upper Quartile	Lower Quartile	-3.21	0.92	0.003*	-5.47	-0.94
		Mean	-1.58	0.81	0.168	-3.58	0.42

* $p < .05$.

4.2.3 Problem Phoneme Improvements

Paired samples t-tests were conducted to investigate improvements in pronunciation accuracy of the 15 problem phonemes identified in the research literature. Tests were conducted based on overall performance, then proficiency and gender levels. The Holm-Bonferroni method was applied to adjust the significance levels of multiple t-test comparisons.

Based on the entire data set, the analysis revealed significant improvement in all phonemes at adjusted alpha levels except for /ð/, /ʃ/, /ʧ/, and /v/, although /ʧ/ reached significance at the non-adjusted level (p < .05). However, mean pre-test scores for the phonemes /ð/, /ʃ/, and /v/ were 95%, 93%, and 91%, respectively, indicating an initial high proficiency level with little scope for improvement. Detailed statistics are provided in Table 5.

Table 5

Paired Samples t-Test Pre-test and Post-test Scores by Phoneme

								95% Confidence Interval	
Phoneme	***MD***	***SD***	***SEM***	***t***	***df***	***p***	**Adjusted alpha**	**Lower Bound**	**Upper Bound**
d	3.52	8.31	0.97	3.62	72	.001*a	0.025	1.58	5.45
ð	0.26	2.35	0.27	0.96	72	.341	0.003	-0.28	0.81
ʤ	9.44	12.11	1.42	6.66	72	.000*a	0.050	6.61	12.26
f	1.51	4.63	0.54	2.78	72	.007*a	0.010	0.43	2.59
g	20.48	11.86	1.39	14.75	72	.000*a	0.050	17.71	23.24
l	2.18	5.06	0.59	3.69	72	.000*a	0.050	1.00	3.36
r	3.69	4.06	0.48	7.76	72	.000*a	0.050	2.74	4.64
s	2.63	5.54	0.65	4.06	72	.000*a	0.050	1.34	3.93
ʃ	0.83	4.52	0.53	1.57	72	.121	0.004	-0.22	1.88
ʧ	2.54	8.09	0.95	2.68	72	.009*	0.007	0.65	4.42
ʊ	40.74	10.42	1.22	33.42	72	.000*a	0.050	38.31	43.17
v	0.82	5.70	0.67	1.24	72	.220	0.004	-0.50	2.15
z	3.75	7.89	0.92	4.06	72	.000*a	0.050	1.90	5.59
ʒ	6.66	10.13	1.19	5.62	72	.000*a	0.050	4.30	9.03
θ	15.70	8.59	1.01	15.61	72	.000*a	0.050	13.69	17.70

* $p < .05$, 2-tailed.

[a] Significant at the adjusted alpha level, 2-tailed.

As Appendix A, Table 3 illustrates, analysis of scores across proficiency levels highlighted a consistent significant improvement at the adjusted alpha levels for four problem phonemes, namely /g/, /r/, /ʊ/, and / θ /. Improvement of the remaining problem phonemes varied by proficiency level.

Students in the lower quartile group reached significant improvement at adjusted alpha levels for the phonemes /ʤ/, /g/, /r/, /s/, /ʊ/, /z/, and / θ /, with /d/, /l/, and /ʒ/ showing significance at non-adjusted levels. Mean quartile group students improved significantly at adjusted levels for /ʤ/, /g/, /r/, /ʊ/, /ʒ/, and / θ /, and significance at non-adjusted levels for /d/, /l/, /s/, and /z/. Students in the upper quartile group reached significant improvement at adjusted levels for

/g/, /r/, /ʊ/, /ʒ/, and /θ/, with /ʤ/ and /f/ reaching significant improvement at non-adjusted levels.

4.2.4 Analysis of the Effect of System Use

Multiple linear regression was used to examine the effect of the four independent frequency-of-use variables (Login, TTS, Evaluation, and Task Review) on overall pronunciation improvement (see Table 6). The model was not significant, (F(4, 68) = 1.36, p = .256, R^2 = .074), indicating that the predictors did not explain the variance in overall improvement. Despite this, the analysis of pre-test and post-test scores revealed significant improvement in pronunciation accuracy, suggesting that while the specific features of ALL-Talk could not be independently attributed to improvement, its overall use may still have positively impacted student performance.

Table 6

Multiple Regression Analysis Predicting Overall Improvement Based on System Usage

Model	*B*	*SE B*	*β*	*t*	*p*	Collinearity Tolerance	VIF
Constant	.251	.665		.378	.707		
Log In	.028	.020	.270	1.418	.161	.375	2.667
TTS/Waveform	.001	.004	.024	.149	.882	.508	1.967
Evaluation	.001	.002	.034	.246	.806	.713	1.404
Task Review	-.069	.038	-.258	-1.819	.073	.678	1.474

4.3 Research question 3: How does gender influence the effectiveness of ASR in counteracting mother tongue pronunciation influences?

As illustrated in Table 2, the Analysis of Covariance (ANCOVA) indicated that gender was not a significant predictor in overall pronunciation accuracy improvement. Nonetheless, the analysis of pre-test and post-test scores highlighted several gender-based disparities.

Descriptive statistics of overall mean scores revealed that males outperformed females on both the pre-test ($t_{\text{Pre-test}}$(71) = 1.03, p = .307, 95% CI [-

.86, 2.71]) and post-test ($t_{Post\text{-}test}(71) = 1.16$, $p = .249$, 95% CI [-.67, 2.52]) (see Table 7). However, as illustrated in Table 3, the mean difference between pre-test and post-test scores reached significance only for females ($p = .018$).

Further, significant improvement was observed for females within the lower and mean quartile groups, but not in the upper quartile group, where female performance declined. A comparison by gender, using Bonferroni post hoc comparisons, revealed that the difference in proficiency levels was significant only among female participants (see Table 4).

Table 7

Mean Scores for Pre-test, Post-test by Gender

	Gender	*N*	*M*	*SD*	*SEM*
Pre-test	Male	16	92.34	3.05	.76
	Female	57	91.42	3.20	.42
Post-test	Male	16	93.21	2.68	.67
	Female	57	92.28	2.86	.38

The results of a paired samples t-test as shown in Appendix A, Table 2 revealed considerable gender performance differences across the problem phonemes. For males, significant improvement was observed at adjusted alpha levels for the phonemes /g/, /r/, /ʊ/, and /θ/, while significant improvements were observed at non-adjusted levels for /d/, /ʤ/, and /ʒ/. In contrast, females reached significant improvement at adjusted alpha levels for the phonemes /d/, /ʤ/, /g/, /l/, /r/, /s/, /ʧ/, /ʊ/, /z/, /ʒ/, and /θ/, while only reaching significant improvement at non-adjusted levels for /f/.

The final level of analysis, which examined the intersection of proficiency levels and gender in relation to phoneme improvement, revealed notable gender differences across groups (see Appendix A, Table 4).

At the lower quartile level, the results revealed that males displayed significant improvement at adjusted alpha levels for the /ʊ/ phoneme while reaching significant improvement at non-adjusted levels for /θ/. In contrast, females displayed significant improvement at adjusted alpha levels for /ʤ/, /g/, /ʊ/, and /θ/, and /l/, /r/, /s/, /z/, and /ʒ/ at non-adjusted levels.

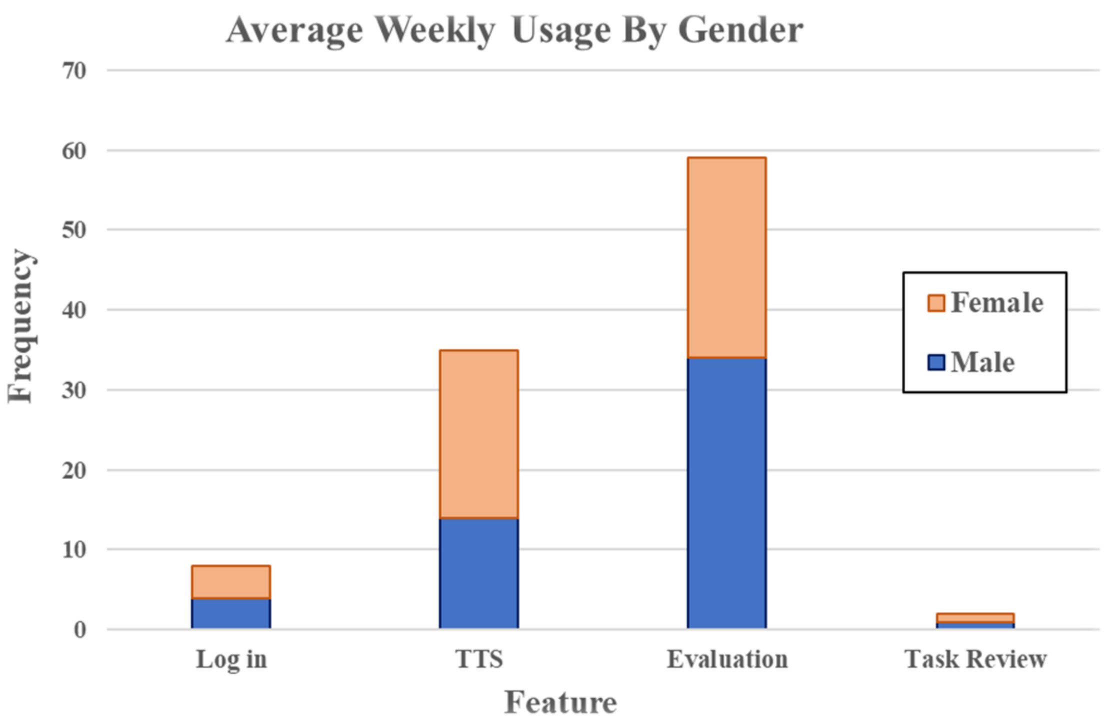
Within the mean quartile group, males reached significant improvement for the phoneme /ʊ/ at adjusted alpha levels and /θ/ at non-adjusted levels. Conversely, females reached significant improvement at adjusted levels for /ʤ/, /g/, /l/, /r/, /ʊ/, /z/, /ʒ/, and /θ/, and at non-adjusted levels for /ð/, /s/, and /ʧ/.

At the upper quartile level, males reached significant improvement at adjusted levels for the phonemes /g/ and /ʊ/and at non-adjusted levels for /ð/, /r/, and /θ/. In contrast, females reached significant improvement at adjusted levels for /g/, /r/, /ʊ/, and /θ/, and /f/ and /ʒ/ at non-adjusted levels.

Descriptive statistics were used to evaluate students' average weekly use of the main features of ALL-Talk, categorized by gender (see Figure 2). The results revealed that, while both genders accessed the system consistently, females used the example speech and waveform feature (TTS) 50% more frequently than males. In contrast, males submitted a higher number of pronunciation attempts for evaluation. Both genders consistently reviewed the evaluation and corrective feedback for their weekly speaking tasks. The greater number of evaluation attempts by males suggests they reattempted the weekly speaking task after reviewing their feedback.

Figure 2

Average Weekly Use of ALL-Talk Main Features By Gender



5. Discussion

This study provided insight into the persistent challenges and progress made in pronunciation among Thai EFL students, revealing nuanced influences of phonetic structure, language interference, and instructional support through technology-enhanced practice.

5.1 Phoneme-specific Pronunciation Challenges and L1 Interference

The analysis of pre- and post-test scores identified consistent challenges in pronouncing specific phonemes, particularly /d/, /ʤ/, /g/, /ʧ/, /ʊ/, /z/, and /θ/. The difficulty with the phoneme /ʊ/, despite its similarity to the Thai monophthong sound ุ found in words like คุณ (pronounced kʊn), is significant. This anomaly suggests that pronunciation issues may arise from subtler coarticulatory influences rather than straightforward differences between L1 and L2 phonetic systems. Additionally, the persistence of mispronunciations with phonemes like /d/, /ʤ/, /s/, /ʧ/, and /z/, all of which are word-final sounds in English but rarely occur as final sounds in Thai, underscores the influence of phonetic context rather than phoneme absence in L1. This highlights the need for pedagogy that addresses both phonetic and contextual factors in pronunciation practice.

5.2 Influence of Markedness and Unmarked Phonemes

The results partially align with Sridhanyarat's (2017) findings on markedness. While some fricatives classified as "marked" (such as /ʃ/, /v/, and /ð/) were found to be easier to pronounce than unmarked sounds like /s/ and /f/, the present study showed that students demonstrated greater initial accuracy with marked fricatives. This finding challenges the MDH, suggesting that exposure and familiarity may play a role in phoneme acquisition beyond markedness alone. Consequently, this underscores the pedagogical value of considering learners' existing proficiencies and introducing sounds based on their relative familiarity and ease, rather than adhering strictly to markedness hierarchies.

5.3 Effectiveness of ALL-Talk for Autonomous Pronunciation Practice

ALL-Talk provided students with a flexible tool for practicing pronunciation autonomously, in line with Haggag (2018) and Behr (2022). Students appreciated the tool's accessibility and privacy, which may counteract the limited pronunciation practice available in traditional classrooms (Thomson & Derwing, 2014). Although specific features were not conclusively linked to pronunciation gains, high engagement with ALL-Talk features suggests that integrating similar tools into EFL pedagogy could encourage students to engage in independent, self-paced learning (Behr, 2022; Olson, 2014; Olson & Offerman, 2021). This supports broader pedagogical strategies promoting autonomous learning, which can be particularly useful in settings with limited class time for pronunciation.

The significant improvements observed in the pronunciation of marked sounds, particularly among students in the lower quartile group, somewhat contrast with Ngo et al. (2023), who suggest that ASR is more effective in EFL contexts for adult learners or those at an intermediate proficiency level. However, Ngo's findings also indicate that visual comparison with native pronunciation models, using tools like waveforms and spectrograms, enhances phonetic awareness and correction—a result that aligns closely with this study and is underscored by Behr (2022) and Olson and Offerman (2021). This suggests that even lower-proficiency learners can benefit from ASR tools that offer visual analysis of speech. Pedagogically, this supports the integration of ASR with pronunciation visuals to aid pronunciation practice for beginner learners, thereby extending ASR's potential as an autonomous learning tool across a wider range of proficiency levels.

5.4 Gender Differences in Pronunciation Gains and Usage Patterns

The study's ANCOVA analysis revealed that gender was not a significant predictor of overall improvement between pre- and post-test scores, aligning with the findings of Jahandar et al. (2012) and Hariri (2012). However, females showed greater improvement within the lower and mean proficiency groups, suggesting that they may have benefited from personalized learning strategies or displayed heightened sensitivity to phonetic nuances, possibly by engaging more actively with TTS and waveform features —a perspective supported by Ngo et al. (2023). This behavior may indicate more effective self-regulation strategies or a greater sensitivity to subtle phonetic discrepancies between L1 and L2 sounds, as noted by Khamkhien (2010).

While males achieved higher mean pre- and post-test scores overall, the gains made by females underscore the potential impact of individualized learning strategies in technology-assisted language learning. These findings support a pedagogical approach that adapts feedback and practice strategies to meet individual learner needs (Alomari, 2024; Ngo et al., 2023; Wiboolyasarin, 2023), highlighting the value of ASR tools like ALL-Talk in accommodating varied learning preferences. Recognizing these gendered engagement patterns could enable instructors to better support diverse learning strategies, particularly within autonomous language-learning environments.

5.5 Implications of Pronunciation Feedback

The algorithm behind ALL-Talk's pronunciation evaluation, which factors in stress, intonation, and completeness, likely contributed to the observed improvements. Females' increased use of TTS and waveform analysis tools may have enabled them to more effectively identify stress patterns and achieve greater intonational accuracy (Spring & Tabuchi, 2022), contributing to their marked improvement. This aligns with Kapranov's (2022) observation that Thai students often substitute /z/ with /s/. In this study, significant gains in /z/ pronunciation following exposure to ALL-Talk's visual feedback suggest that the technology helped participants distinguish phonetic contrasts more accurately. These findings have practical implications for pronunciation pedagogy, highlighting the value of integrating visual and auditory feedback to support students in independently refining their pronunciation skills.

5.6 Summary

The study highlights that pronunciation challenges for Thai EFL learners are influenced by complex interactions between L1 and L2 phonetic structures, markedness of phonemes, and contextual factors, and underscores the need to consider complex L1-L2 phonetic and contextual interactions in pronunciation teaching. ALL-Talk emerged as a beneficial tool for autonomous practice, particularly for female students, although gender was not a significant predictor of overall improvement. Future research could explore the longitudinal effects of similar interventions to further examine the interplay of gender, learning strategies, and technology-driven language acquisition outcomes.

6. Limitations and Future Research

While this study focused on the pronunciation of known problem phonemes, it overlooked problems relating to consonant clusters, a known issue for Thai learners (Jahandar et al., 2012; Lai et al., 2009). Furthermore, it did not specifically analyze the influence of phoneme position, specifically word-ending sounds. Future research should explore ASR applications that address consonant clusters and phoneme positioning to provide a more comprehensive understanding of pronunciation challenges. Additionally, this study's relatively small sample size limits the generalizability of its findings, particularly regarding gender patterns, suggesting that larger studies are necessary to validate these results.

7. Conclusion

Students in this study overcame pronunciation barriers where the L2 sound was absent in the L1 but appeared to have had greater difficulty pronouncing English phonemes where there were differences in the L1 and L2 phonological contexts, such as syllable and word ending sounds. The persistence of pronunciation difficulties with phonemes like /ʊ/ and /ʧ/, despite their presence in Thai, underscores the importance of considering phonotactic constraints in instructional approaches.

This study advocates for integrating ASR tools into language learning to support autonomous, personalized feedback, which may be especially effective for EFL learners who have limited classroom opportunities to focus on pronunciation. The findings also suggest that gender-specific learning strategies may play a role in how learners engage with feedback, underscoring the value of tailoring ASR feedback to support diverse learning preferences.

8. About the Author

Simon Moxon is an EFL lecturer at Walailak University, Nakhon Si Thammarat, Thailand. He holds a first-class honors degree in Applied Computing from Staffordshire University, UK, as well as a master's degree and PhD in Teaching and Technology from Assumption University, Thailand. His research interests include pronunciation and CALL.

9. Acknowledgements

The author would like to thank Walailak University for funding this study (Grant Number: WU66264) and the Human Research Ethics Committee of Walailak University for considering and approving this research (Approval Number: WUEC-

23-208-01). The author declares that there are no conflicts of interest regarding the publication of this article.

10. References

Afshari, S. S., & Ketabi, S. (2017). Current trends and future directions in teaching English pronunciation. *International Journal of Research Studies in Language Learning*, *5*(3), 83–91. https://doi.org/10.5861/ijrsll.2016.1437

Alomari, N. M. (2024). Action research in ESL: Enhancing English oral proficiency in an asynchronous speech course. *Journal of Language Teaching and Research*, *15*(1), 1–8. https://doi.org/10.17507/jltr.1501.01

Becker, B. P. (1995). *Thai for beginners*. Paiboon Publishing.

Behr, N. S. (2022). English diphthong characteristics produced by Thai EFL learners: Individual practice using PRAAT. *Computer Assisted Language Learning Electronic Journal, 23*(1), 401–424.

Brett, D. (2004). Computer generated feedback on vowel production by learners of English as a second language. *ReCALL, 16*(1), 103–113. https://doi.org/10.1017/S0958344004000813

Bryla-Cruz, A. (2021). The gender factor in the perception of English segments by non-native speakers. *Studies in Second Language Learning and Teaching, 11*(1), 103–131. https://doi.org/10.14746/ssllt.2021.11.1.5

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27*(2), 315–330. https://doi.org/10.1111/j.1467-1770.1977.tb00124.x

Fauzi, I. (2021). The variability in phonology of Indonesian learner's interlanguage: A case study on English marked-fricatives. *International Journal of Language Education, 5*(4), 285–295. https://doi.org/10.26858/ijole.v5i4.19468

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). Sage.

Fitriani, N., & Zulkarnain, S. I. (2019). An investigation of mother tongue influence on EFL learners during their speaking performance. *Southeast Asia Language Teaching and Learning Journal, 2*(2), 30–35. https://doi.org/10.35307/saltel.v2i2.29

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Gabriel, R. (2023). Noises and colors: Two untraditional methods of teaching ESL/ EFL pronunciation. *Journal of Language Teaching and Research, 14*(4), 882–892. https://doi.org/10.17507/jltr.1404.04

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism, 10*(2), 486–489. https://doi.org/10.5812/ijem.3505

Haggag, H. M. (2018). Teaching phonetics using a mobile-based application in an EFL context. *European Scientific Journal, 14*(4), 189–204. https://doi.org/10.19044/esj.2018.v14n14p189

Hariri, M. (2012). A review of literature: A gender-based study of pronunciation accuracy. *Research Journal of Applied Sciences, Engineering and Technology, 4*(22), 4861–4864. https://maxwellsci.com/print/rjaset/v4-4861-4864.pdf

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70. http://www.jstor.org/stable/4615733

Huang, X., & Jia, X. (2016). Corrective feedback on pronunciation: Students' and teachers' perceptions. *International Journal of English Linguistics, 6*(6), 245–254. https://doi.org/10.5539/ijel.v6n6p245

Isarankura, S. (2015). Using the audio-articulation method to improve EFL learners' pronunciation of the English /v/ sound. *Thammasat Review, 18*(2), 116–137. https://sc01.tci-thaijo.org/index.php/tureview/article/view/67878/55358

Jahandar, S., Khodabandehlou, M., Seyedi, G., & Abadi, R. M. D. (2012). A gender-based approach to pronunciation accuracy of advanced EFL learners. *International Journal of Scientific & Engineering Research*, *3*(6), 1–9. https://www.ijser.org/researchpaper/A-Gender-based-Approach-to-Pronunciation-Accuracy-of-Advanced-EFL-Learners.pdf

Kanokpermpoon, M. (2007). Thai and English consonantal sounds: A problem or a potential for EFL learning? *ABAC Journal, 27*(1), 57–66. http://www.assumptionjournal.au.edu/index.php/abacjournal/article/view/583/520

Kapranov, O. (2022). The English fricative consonant /z/ as a challenge to Norwegian L1 EFL learners: An error analysis of phonemic transcriptions. *Philologia Estonica Tallinnensis, 7*, 148–184. https://doi.org/10.22601/pet.2022.07.06

Kayaoğlu, M. N. (2019). Rehabilitating fossilized consonants through computer-aided and animated material. *Hacettepe University Journal of Education, 34*(2), 460–472. https://doi.org/10.16986/HUJE.2018040668

Khamkhien, A. (2010). Thai learners' English pronunciation competence: Lesson learned from word stress assignment. *Journal of Language Teaching and Research, 1*(6), 757–764. https://doi.org/10.4304/jltr.1.6.757-764

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press.

Lai, Y.-S., Tsai, H.-H., & Yu, P.-T. (2009). A multimedia English learning system using HMMs to improve phonemic awareness for English learning. *Educational Technology & Society, 12*(3), 266–281. http://doi.org/10.1109/MUE.2007.30

Mahzoun, Z., & Han, T. (2019). The effects of consonant phonemes' position across the word on pronunciation errors: An empirical study of Turkish EFL learners. *The Reading Matrix: An International Online Journal, 19*(2), 21–32. https://www.readingmatrix.com/files/21-632icisq.pdf

Mantei, J., Kervin, L., & Jones, P. (2021). Examining pedagogies for teaching phonics: Lessons from early childhood classrooms. *The Australian*

Educational Researcher, *49*(4), 743–760. https://doi.org/10.1007/s13384-021-00454-8

Moxon, S. (2021). Exploring the effects of automated pronunciation evaluation on L2 students in Thailand. *IAFOR Journal of Education: Language Learning in Education, 9*(3), 41–56. https://doi.org/10.22492/ije.9.3.03

Moxon, S. (2023). Pronunciation Coach 3D. *Computer Assisted Language Learning Electronic Journal, 24*(1), 205–221. https://old.callej.org/journal/24-1/Moxon2023.pdf

Moxon, S. (2024). ALL-Talk: Enhancing EFL pronunciation with Microsoft Azure speech services. *ABAC Journal, 44*(4), 139–161. https://doi.org/10.59865/abacj.2024.58

Ngo, T. T. -N., Chen, H. H. -J., & Lai, K. K. -W. (2023). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, *36*(1), 4–21. https://doi.org/10.1017/S0958344023000113

Nushi, M., & Sadeghi, M. (2021). A critical review of ELSA: A pronunciation app. *Computer Assisted Language Learning Electronic Journal, 22*(3), 287–302.

Olson, D. J. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania, 97*(1), 47–68. https://doi.org/10.1353/hpn.2014.0030

Olson, D. J. & Offerman, H. M. (2021). Maximizing the effect of visual feedback for pronunciation instruction: A comparative analysis of three approaches. *Journal of Second Language Pronunciation, 7*(1), 89–115. https://doi.org/10.1075/jslp.20005.ols

Osborne, J. W., & Overbay, A. (2004) The power of outliers (and why researchers should ALWAYS check for them). *Practical Assessment, Research, and Evaluation, 9*(6), 1–8. https://doi.org/10.7275/qf69-7k43

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). Thomson Learning.

Qader, M., Rahman, M., & Monira, S. (2023). Application of phonetics and phonology in secondary level education for reducing Bengali interference in English pronunciation in Bangladesh. *World Journal of English Language, 13*(7), 597–612. http://doi.org/10.5430/wjel.v13n7p597

Setter, J., & Jenkins, J. (2005). State-of-the-art review article: Pronunciation. *Language Teaching, 38*(1), 1–17. https://doi.org/10.1017/S026144480500251X

Spring, R., & Tabuchi, R. (2022). The role of ASR training in EFL pronunciation improvement: An in-depth look at the impact of treatment length and guided practice on specific points. *Computer Assisted Language Learning Electronic Journal, 23*(3), 163–185. https://old.callej.org/journal/23-3/Spring-Tabuchi2022.pdf

Sridhanyarat, K. (2017). The acquisition of L2 fricatives in Thai learners' interlanguage. *The Southeast Asian Journal of English Language Studies, 23*(1), 15–34. http://doi.org/10.17576/3L-2017-2301-02

Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. *Frontiers in Psychology*, *14*, Article 1210187. https://doi.org/10.3389/fpsyg.2023.1210187

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36*(3), 326–344. https://doi.org/10.1093/applin/amu076

Tingsabadh, M., & Abramson, A. (1993). Thai. *Journal of the International Phonetic Association, 23*(1), 24–28. http://doi.org/10.1017/S0025100300004746

van den Doel, R., Pinget, A.-F. C. H., & Quené, H. (2018). Non-native attitudes to / θ / and /ð/: A European case study. *Research in Language, 16*(4), 407–427. https://doi.org/10.2478/rela-2018-0020

Wiboolyasarin, W., Tiranant, P, Khumsat, T, Ngamnikorn, T., Wiboolysarin, K., Korbuakaew, S., & Jinowat, N. (2023). Preferences for oral corrective feedback: Are language proficiency, first language, foreign language classroom anxiety, and enjoyment involved? *Journal of Language and Education, 9*(1), 172–184. https://doi.org/10.17323/jle.2023.16141

Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *Sage Open, 13*(3), 1–8. https://doi.org/10.1177/21582440231193818

11. Appendices

11.1 Appendix A

Table 1

Descriptive Statistics for Pre-Test Phoneme Sounds by Proficiency Level and Gender (Overall Mean Scores Below 75%)

Phoneme	*Proficiency Level*	*Gender*	*N*	Min	Max	*M*	*SD*
d	All	Both	73	54.70	93.00	73.23	8.94
	Lower Quartile	Male	4	57.40	78.60	69.73	9.29
		Female	15	54.70	84.70	66.45	7.76
	Mean	Male	6	55.90	84.50	72.15	11.41
		Female	29	60.40	85.40	72.27	6.46
	Upper Quartile	Male	6	72.30	83.70	77.98	4.72
		Female	13	70.60	93.00	82.79	7.23
ʤ	All	Both	73	31.8	95.9	68.10	13.99
	Lower Quartile	Male	4	45.60	76.60	65.43	13.68
		Female	15	39.40	75.80	56.18	10.14
	Mean	Male	6	45.70	87.10	69.58	14.31
		Female	29	31.80	90.20	65.98	12.74
	Upper Quartile	Male	6	71.50	87.50	81.58	6.74
		Female	13	60.70	95.90	80.52	8.68
g	All	Both	73	34.40	86.00	62.15	10.60
	Lower Quartile	Male	4	55.70	81.30	69.75	12.86
		Female	15	34.40	86.00	58.91	14.33
	Mean	Male	6	62.60	77.60	68.28	5.40
		Female	29	43.00	78.90	59.46	9.68
	Upper Quartile	Male	6	52.40	73.00	63.43	6.96
		Female	13	56.00	78.70	66.14	7.42
ʧ	All	Both	73	50.80	95.4	73.97	9.98

Phoneme	*Proficiency Level*	*Gender*	*N*	Min	Max	*M*	*SD*
	Lower Quartile	Male	4	63.10	73.40	68.50	4.90
		Female	15	50.80	81.50	64.93	9.42
	Mean	Male	6	74.40	83.90	77.90	3.75
		Female	29	57.00	90.90	73.53	7.92
	Upper Quartile	Male	6	76.90	95.40	86.12	6.03
		Female	13	64.30	94.60	79.67	9.55
ʊ	All	Both	73	40.8	76.6	53.63	7.50
	Lower Quartile	Male	4	50.30	61.80	57.45	5.32
		Female	15	44.50	71.70	53.97	7.81
	Mean	Male	6	45.10	76.60	58.45	11.66
		Female	29	40.80	66.50	52.79	6.58
	Upper Quartile	Male	6	44.00	67.20	51.75	8.19
		Female	13	41.70	70.80	52.59	7.23
z	All	Both	73	49.8	95.1	73.16	12.22
	Lower Quartile	Male	4	57.90	76.20	66.98	10.37
		Female	15	52.40	81.90	62.36	7.49
	Mean	Male	6	49.80	81.50	68.15	13.27
		Female	29	50.50	86.20	71.93	9.77
	Upper Quartile	Male	6	70.70	90.60	83.35	8.39
		Female	13	79.50	95.10	87.85	4.22
θ	All	Both	73	44.30	88.10	69.54	8.63
	Lower Quartile	Male	4	68.50	73.70	71.65	2.39
		Female	15	51.60	73.00	61.64	6.68
	Mean	Male	6	63.40	80.30	68.52	6.28
		Female	29	44.30	86.10	68.93	8.23
	Upper Quartile	Male	6	71.60	80.10	76.35	3.24
		Female	13	64.90	88.10	76.71	7.37

Table 2

Paired Samples t Test Pre-test and Post-test Scores by Phoneme and Gender

									95% Confidence Interval	
								Adjusted	Lower	Upper
Phoneme	**Gender**	***MD***	***SD***	***SEM***	***t***	***df***	***p***	**alpha**	**Bound**	**Bound**
d	Male	5.04	8.31	2.08	2.43	15	.028*	0.005	0.61	9.46
	Female	3.09	8.33	1.10	2.80	56	.007*a	0.008	0.88	5.30
ð	Male	0.19	2.18	0.54	0.34	15	.735	0.001	-0.97	1.35
	Female	0.28	2.41	0.32	0.89	56	.377	0.002	-0.36	0.92
ʤ	Male	6.69	9.71	2.43	2.76	15	.015*	0.007	1.52	11.87
	Female	10.21	12.67	1.68	6.08	56	.000*a	0.050	6.85	13.57
f	Male	1.88	4.08	1.02	1.84	15	.085	0.003	-0.29	4.06
	Female	1.40	4.80	0.64	2.21	56	.032*	0.004	0.13	2.68
g	Male	18.52	13.93	3.48	5.32	15	.000*a	0.050	11.10	25.94
	Female	21.03	11.30	1.50	14.05	56	.000*a	0.050	18.03	24.02
l	Male	1.01	6.80	1.70	0.60	15	.560	0.001	-2.61	4.64
	Female	2.51	4.47	0.59	4.24	56	.000*a	0.050	1.33	3.70
r	Male	3.47	3.87	0.97	3.58	15	.003*a	0.013	1.41	5.53
	Female	3.76	4.15	0.55	6.84	56	.000*a	0.050	2.66	4.86
s	Male	2.34	6.03	1.51	1.56	15	.141	0.002	-0.87	5.55
	Female	2.72	5.45	0.72	3.76	56	.000*a	0.050	1.27	4.16
ʃ	Male	1.49	3.54	0.88	1.68	15	.113	0.003	-0.40	3.37
	Female	0.64	4.76	0.63	1.02	56	.312	0.002	-0.62	1.91
ʧ	Male	-1.89	5.70	1.42	-1.33	15	.204	0.002	-4.93	1.14
	Female	3.78	8.26	1.09	3.46	56	.001*a	0.025	1.59	5.97
ʊ	Male	39.23	12.57	3.14	12.48	15	.000*a	0.050	32.53	45.93
	Female	41.16	9.81	1.30	31.67	56	.000*a	0.050	38.55	43.76
v	Male	1.39	6.46	1.61	0.86	15	.403	0.001	-2.05	4.83
	Female	0.67	5.52	0.73	0.91	56	.365	0.002	-0.80	2.13
z	Male	2.77	9.25	2.31	1.20	15	.250	0.002	-2.16	7.70
	Female	4.02	7.53	1.00	4.03	56	.000*a	0.050	2.02	6.02
ʒ	Male	5.05	9.06	2.27	2.23	15	.042*	0.004	0.22	9.88
	Female	7.11	10.44	1.38	5.14	56	.000*a	0.050	4.34	9.88
θ	Male	13.62	7.22	1.80	7.55	15	.000*a	0.050	9.77	17.47
	Female	16.28	8.91	1.18	13.80	56	.000*a	0.050	13.92	18.65

[*] $p < .05$, 2-tailed.

[a] Significant at the adjusted alpha level, 2-tailed.

Table 3

Paired Samples t-Test Pre-test and Post-test Scores by Phoneme and Proficiency Level

Phoneme	Proficiency Level	*MD*	*SD*	*SEM*	*t*	*df*	*p*	Adjusted Alpha	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
d	Lower Quartile	4.75	8.39	1.92	2.47	18	.024*	0.003	0.70	8.79
	Mean	3.03	8.63	1.46	2.08	34	.045*	0.002	0.07	5.99
	Upper Quartile	3.18	7.93	1.82	1.75	18	.098	0.002	-0.64	7.00
ð	Lower Quartile	-0.03	2.80	0.64	-0.04	18	.968	0.001	-1.37	1.32
	Mean	0.64	2.41	0.41	1.57	34	.125	0.001	-0.19	1.47
	Upper Quartile	-0.14	1.66	0.38	-0.37	18	.713	0.001	-0.94	0.66
ʤ	Lower Quartile	10.84	13.70	3.14	3.45	18	.003*a	0.013	4.23	17.44
	Mean	10.68	12.60	2.13	5.01	34	.000*a	0.050	6.35	15.01
	Upper Quartile	5.75	8.92	2.05	2.81	18	.012*	0.005	1.45	10.05
f	Lower Quartile	2.28	6.42	1.47	1.55	18	.139	0.001	-0.81	5.38
	Mean	0.69	4.07	0.69	1.00	34	.325	0.001	-0.71	2.08
	Upper Quartile	2.25	3.29	0.76	2.97	18	.008*	0.006	0.66	3.83
g	Lower Quartile	19.73	15.52	3.56	5.54	18	.000*a	0.050	12.25	27.21
	Mean	20.19	11.23	1.90	10.64	34	.000*a	0.050	16.33	24.04
	Upper Quartile	21.76	9.02	2.07	10.51	18	.000*a	0.050	17.41	26.11
l	Lower Quartile	3.43	6.47	1.48	2.31	18	.033*	0.003	0.31	6.55
	Mean	2.32	4.90	0.83	2.80	34	.008*	0.006	0.64	4.01
	Upper Quartile	0.68	3.31	0.76	0.89	18	.384	0.001	-0.92	2.28
r	Lower Quartile	3.67	5.01	1.15	3.20	18	.005*a	0.010	1.26	6.09
	Mean	4.05	4.16	0.70	5.76	34	.000*a	0.050	2.62	5.48
	Upper Quartile	3.06	2.77	0.63	4.82	18	.000*a	0.050	1.72	4.39
s	Lower Quartile	4.82	6.56	1.51	3.20	18	.005*a	0.010	1.65	7.98
	Mean	1.98	5.34	0.90	2.19	34	.035*	0.003	0.14	3.81
	Upper Quartile	1.66	4.36	1.00	1.66	18	.114	0.002	-0.44	3.77
ʃ	Lower Quartile	1.57	5.16	1.18	1.33	18	.202	0.001	-0.92	4.05
	Mean	0.55	4.92	0.83	0.66	34	.514	0.001	-1.14	2.24
	Upper Quartile	0.61	2.92	0.67	0.90	18	.378	0.001	-0.80	2.01
ʧ	Lower Quartile	4.92	11.08	2.54	1.93	18	.069	0.002	-0.42	10.26
	Mean	2.19	7.31	1.24	1.77	34	.086	0.002	-0.33	4.70
	Upper Quartile	0.81	5.35	1.23	0.66	18	.520	0.001	-1.77	3.38
ʊ	Lower Quartile	35.95	11.11	2.55	14.10	18	.000*a	0.050	30.60	41.31
	Mean	40.46	10.54	1.78	22.70	34	.000*a	0.050	36.84	44.08
	Upper Quartile	46.02	6.83	1.57	29.37	18	.000*a	0.050	42.73	49.31

Phoneme	Proficiency Level	*MD*	*SD*	*SEM*	*t*	*df*	*p*	Adjusted Alpha	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
v	Lower Quartile	1.15	5.42	1.24	0.92	18	.368	0.001	-1.47	3.76
	Mean	1.07	6.58	1.11	0.96	34	.345	0.001	-1.20	3.33
	Upper Quartile	0.06	4.20	0.96	0.06	18	.953	0.001	-1.96	2.08
z	Lower Quartile	6.77	9.62	2.21	3.07	18	.007*a	0.007	2.14	11.41
	Mean	3.40	7.60	1.28	2.64	34	.012*	0.004	0.79	6.01
	Upper Quartile	1.36	5.59	1.28	1.06	18	.303	0.001	-1.33	4.05
ʒ	Lower Quartile	7.38	11.28	2.59	2.85	18	.011*	0.005	1.94	12.82
	Mean	6.49	10.41	1.76	3.69	34	.001*a	0.025	2.91	10.07
	Upper Quartile	6.26	8.82	2.02	3.10	18	.006*a	0.008	2.01	10.51
θ	Lower Quartile	16.95	9.42	2.16	7.84	18	.000*a	0.050	12.41	21.49
	Mean	16.30	8.27	1.40	11.67	34	.000*a	0.050	13.46	19.14
	Upper Quartile	13.34	8.31	1.91	6.99	18	.000*a	0.050	9.33	17.34

[*] $p < .05$, 2-tailed.

[a] Significant at the adjusted alpha level, 2-tailed.

Table 4

Paired Samples t-Test Pre-test and Post-test Scores by Phoneme, Proficiency Level, and Gender

Phoneme	Proficiency Level	Gender	*MD*	*SD*	*SEM*	*t*	*df*	*p*	Adjusted Alpha	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
d	Lower Quartile	Male	4.72	7.94	3.97	1.19	3	.319	0.001	-7.90	17.35
		Female	4.75	8.78	2.27	2.10	14	.055	0.001	-0.11	9.61
	Mean	Male	2.83	10.28	4.20	0.68	5	.529	0.000	-7.95	13.62
		Female	3.07	8.45	1.57	1.96	28	.060	0.001	-0.14	6.29
	Upper Quartile	Male	7.45	7.13	2.91	2.56	5	.051	0.001	-0.03	14.93
		Female	1.21	7.74	2.15	0.56	12	.584	0.000	-3.47	5.88
ð	Lower Quartile	Male	0.93	2.97	1.49	0.62	3	.578	0.000	-3.80	5.65
		Female	-0.28	2.80	0.72	-0.39	14	.704	0.000	-1.83	1.27
	Mean	Male	-1.10	2.28	0.93	-1.18	5	.291	0.001	-3.49	1.29
		Female	1.00	2.31	0.43	2.34	28	.027*	0.002	0.12	1.88
	Upper Quartile	Male	0.98	0.75	0.30	3.23	5	.023*	0.002	0.20	1.77
		Female	-0.66	1.72	0.48	-1.39	12	.191	0.001	-1.70	0.38

Phoneme	Proficiency Level	Gender	*MD*	*SD*	*SEM*	*t*	*df*	*p*	Adjusted Alpha	95% Confidence Interval Lower Bound	Upper Bound
ʤ	Lower Quartile	Male	4.90	14.23	7.11	0.69	3	.540	0.000	-17.74	27.54
		Female	12.42	13.61	3.51	3.53	14	.003*a	0.013	4.88	19.96
	Mean	Male	7.30	8.79	3.59	2.04	5	.097	0.001	-1.92	16.52
		Female	11.38	13.27	2.46	4.62	28	.000*a	0.050	6.33	16.42
	Upper Quartile	Male	7.28	8.97	3.66	1.99	5	.103	0.001	-2.13	16.70
		Female	5.05	9.17	2.54	1.98	12	.071	0.001	-0.49	10.59
f	Lower Quartile	Male	4.33	5.77	2.89	1.50	3	.231	0.001	-4.86	13.51
		Female	1.74	6.66	1.72	1.01	14	.329	0.001	-1.95	5.43
	Mean	Male	1.28	4.53	1.85	0.69	5	.519	0.000	-3.47	6.04
		Female	0.56	4.04	0.75	0.75	28	.460	0.000	-0.97	2.10
	Upper Quartile	Male	0.85	1.73	0.71	1.21	5	.282	0.001	-0.96	2.66
		Female	2.89	3.69	1.02	2.83	12	.015*	0.003	0.66	5.12
g	Lower Quartile	Male	14.40	11.44	5.72	2.52	3	.086	0.001	-3.81	32.61
		Female	21.15	16.48	4.25	4.97	14	.000*a	0.050	12.03	30.28
	Mean	Male	12.15	16.61	6.78	1.79	5	.133	0.001	-5.29	29.59
		Female	21.85	9.33	1.73	12.61	28	.000*a	0.050	18.30	25.40
	Upper Quartile	Male	27.63	8.08	3.30	8.38	5	.000*a	0.050	19.16	36.11
		Female	19.05	8.35	2.32	8.22	12	.000*a	0.050	14.00	24.09
l	Lower Quartile	Male	1.18	6.25	2.55	0.46	5	.662	0.000	-5.38	7.74
		Female	3.49	5.39	1.39	2.51	14	.025*	0.002	0.51	6.48
	Mean	Male	-0.62	4.84	1.97	-0.31	5	.767	0.000	-5.69	4.46
		Female	2.56	4.67	0.87	2.95	28	.006*a	0.007	0.78	4.34
	Upper Quartile	Male	3.20	10.74	5.37	0.60	3	.593	0.000	-13.89	20.29
		Female	1.28	2.34	0.65	1.96	12	.073	0.001	-0.14	2.69
r	Lower Quartile	Male	2.40	4.91	2.00	1.20	5	.285	0.001	-2.75	7.55
		Female	3.29	5.44	1.40	2.34	14	.034*	0.002	0.28	6.31
	Mean	Male	3.45	3.45	1.41	2.45	5	.058	0.001	-0.17	7.07
		Female	4.39	4.00	0.74	5.91	28	.000*a	0.050	2.87	5.91
	Upper Quartile	Male	5.10	3.03	1.52	3.37	3	.044*	0.002	0.28	9.92
		Female	2.88	2.53	0.70	4.10	12	.001*a	0.025	1.35	4.41
s	Lower Quartile	Male	-1.07	4.86	1.98	-0.54	5	.614	0.000	-6.17	4.04
		Female	4.37	6.42	1.66	2.63	14	.020*	0.003	0.81	7.92
	Mean	Male	2.98	4.51	1.84	1.62	5	.166	0.001	-1.75	7.72
		Female	2.61	5.29	0.98	2.65	28	.013*	0.004	0.59	4.62
	Upper Quartile	Male	6.50	7.81	3.91	1.66	3	.195	0.001	-5.93	18.93
		Female	1.05	4.33	1.20	0.88	12	.398	0.000	-1.56	3.67
ʃ	Lower Quartile	Male	-0.67	2.43	0.99	-0.67	5	.531	0.000	-3.22	1.88

Phoneme	Proficiency Level	Gender	*MD*	*SD*	*SEM*	*t*	*df*	*p*	Adjusted Alpha	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
		Female	1.02	5.44	1.41	0.73	14	.480	0.000	-1.99	4.03
	Mean	Male	2.22	3.65	1.49	1.49	5	.197	0.001	-1.61	6.05
		Female	0.80	5.28	0.98	0.82	28	.422	0.000	-1.21	2.81
	Upper Quartile	Male	3.63	3.76	1.88	1.93	3	.150	0.001	-2.36	9.61
		Female	-0.14	2.31	0.64	-0.22	12	.832	0.000	-1.53	1.26
ʧ	Lower Quartile	Male	-3.40	6.07	2.48	-1.37	5	.228	0.001	-9.77	2.97
		Female	5.82	12.11	3.13	1.86	14	.084	0.001	-0.89	12.53
	Mean	Male	-2.67	5.32	2.17	-1.23	5	.274	0.001	-8.25	2.91
		Female	3.34	7.09	1.32	2.54	28	.017*	0.003	0.64	6.04
	Upper Quartile	Male	1.53	5.71	2.85	0.53	3	.630	0.000	-7.56	10.61
		Female	2.41	4.73	1.31	1.84	12	.091	0.001	-0.45	5.26
ʊ	Lower Quartile	Male	35.60	15.02	6.13	5.80	5	.002*a	0.017	19.84	51.36
		Female	36.93	11.76	3.04	12.16	14	.000*a	0.050	30.42	43.45
	Mean	Male	47.50	8.29	3.39	14.03	5	.000*a	0.050	38.80	56.20
		Female	41.47	9.41	1.75	23.73	28	.000*a	0.050	37.89	45.05
	Upper Quartile	Male	32.28	8.50	4.25	7.59	3	.005*a	0.008	18.74	45.81
		Female	45.34	6.30	1.75	25.93	12	.000*a	0.050	41.53	49.15
v	Lower Quartile	Male	0.58	8.99	3.67	0.16	5	.880	0.000	-8.85	10.01
		Female	0.61	5.47	1.41	0.43	14	.674	0.000	-2.42	3.63
	Mean	Male	1.00	4.78	1.95	0.51	5	.630	0.000	-4.02	6.02
		Female	1.17	6.18	1.15	1.02	28	.318	0.001	-1.18	3.52
	Upper Quartile	Male	3.18	5.46	2.73	1.16	3	.329	0.001	-5.52	11.87
		Female	-0.38	4.03	1.12	-0.34	12	.742	0.000	-2.81	2.06
z	Lower Quartile	Male	0.90	10.89	4.44	0.20	5	.848	0.000	-10.53	12.33
		Female	7.15	9.71	2.51	2.85	14	.013*	0.004	1.78	12.53
	Mean	Male	2.92	7.84	3.20	0.91	5	.404	0.000	-5.31	11.15
		Female	3.91	6.88	1.28	3.06	28	.005*a	0.008	1.30	6.53
	Upper Quartile	Male	5.35	10.59	5.29	1.01	3	.387	0.000	-11.50	22.20
		Female	0.64	4.41	1.22	0.52	12	.611	0.000	-2.02	3.30
ʒ	Lower Quartile	Male	3.13	6.88	2.81	1.12	5	.315	0.001	-4.08	10.35
		Female	7.92	11.37	2.94	2.70	14	.017*	0.003	1.62	14.22
	Mean	Male	6.77	9.97	4.07	1.66	5	.157	0.001	-3.70	17.23
		Female	7.18	10.97	2.04	3.53	28	.001*a	0.025	3.01	11.35
	Upper Quartile	Male	5.35	12.36	6.18	0.87	3	.450	0.000	-14.32	25.02
		Female	6.03	8.66	2.40	2.51	12	.027*	0.002	0.80	11.27
θ	Lower Quartile	Male	13.05	7.45	3.04	4.29	5	.008*	0.006	5.23	20.87
		Female	17.62	10.36	2.68	6.59	14	.000*a	0.050	11.88	23.36
	Mean	Male	13.65	9.34	3.81	3.58	5	.016*	0.003	3.84	23.46

Phoneme	Proficiency Level	Gender	*MD*	*SD*	*SEM*	*t*	*df*	*p*	Adjusted Alpha	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
		Female	16.98	8.39	1.56	10.90	28	.000*a	0.050	13.79	20.17
	Upper Quartile	Male	14.43	4.58	2.29	6.30	3	.008*	0.006	7.14	21.71
		Female	13.19	8.20	2.27	5.80	12	.000*a	0.050	8.24	18.15

[*] $p < .05$, 2-tailed.

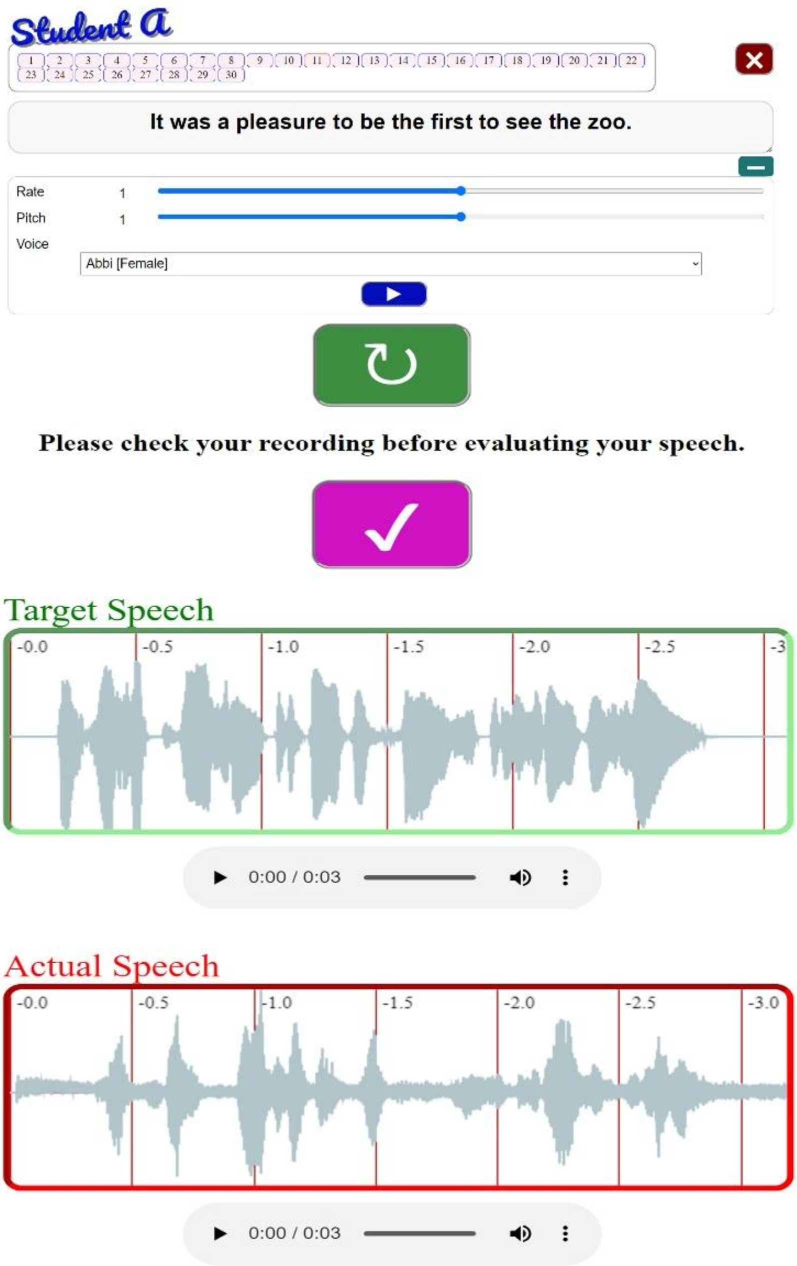
[a] Significant at the adjusted alpha level, 2-tailed.

[**] *t* test not calculated for Phoneme j, for Males in the Upper Quartile group due to zero variance in scores as both pre-test and post-test results were 100%.

11.2 Appendix B

Figure 1

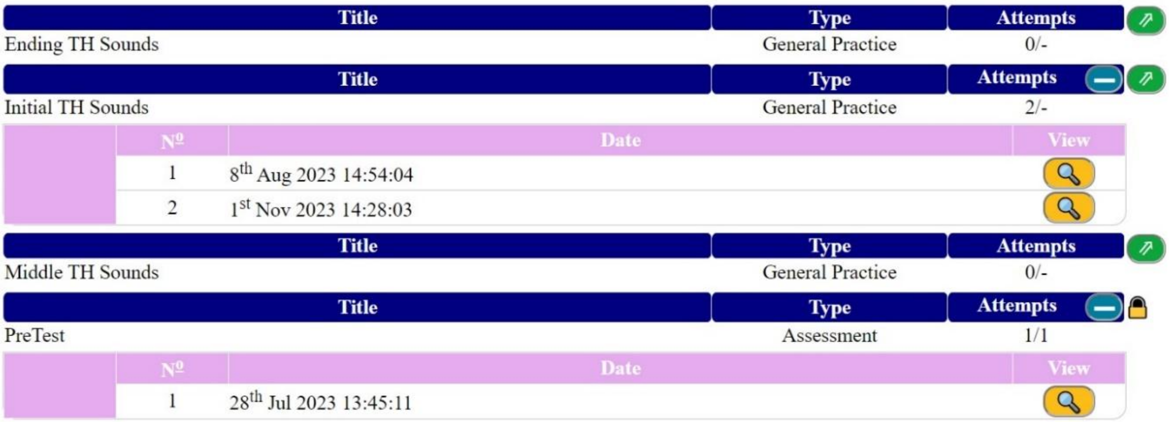
Example Speaking Task Showing TTS and Waveform Feedback



Note. The creation of the example speech is optional. However, it serves as a means for the participant to compare their speech audibly and visually through the use of audio playback and waveforms.

Figure 2

Speaking Task List and Attempt History



Title	Type	Attempts
Ending TH Sounds	General Practice	0/-
Title	**Type**	**Attempts**
Initial TH Sounds	General Practice	2/-

№	Date	View
1	8th Aug 2023 14:54:04	
2	1st Nov 2023 14:28:03	

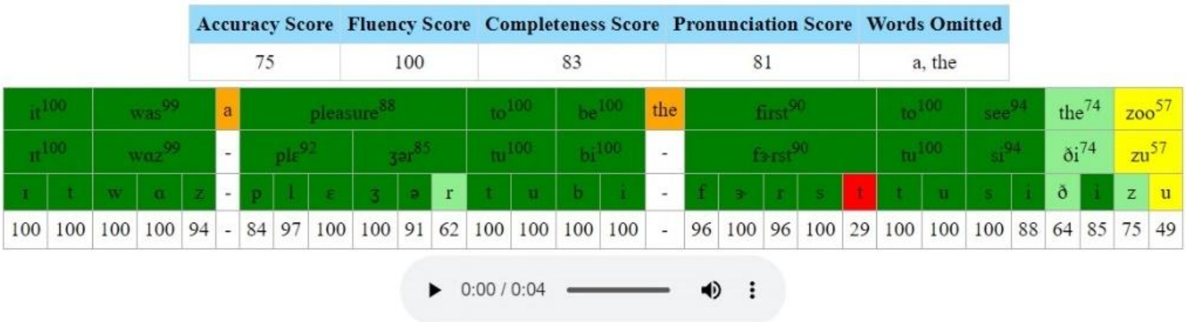
Title	Type	Attempts
Middle TH Sounds	General Practice	0/-
Title	**Type**	**Attempts**
PreTest	Assessment	1/1

№	Date	View
1	28th Jul 2023 13:45:11	

Note. All speaking task attempts are stored chronologically against their corresponding task and may be reviewed at any time. The illustration shows available, attempted, and locked tasks.

Figure 3

Speaking Task Evaluation Scores



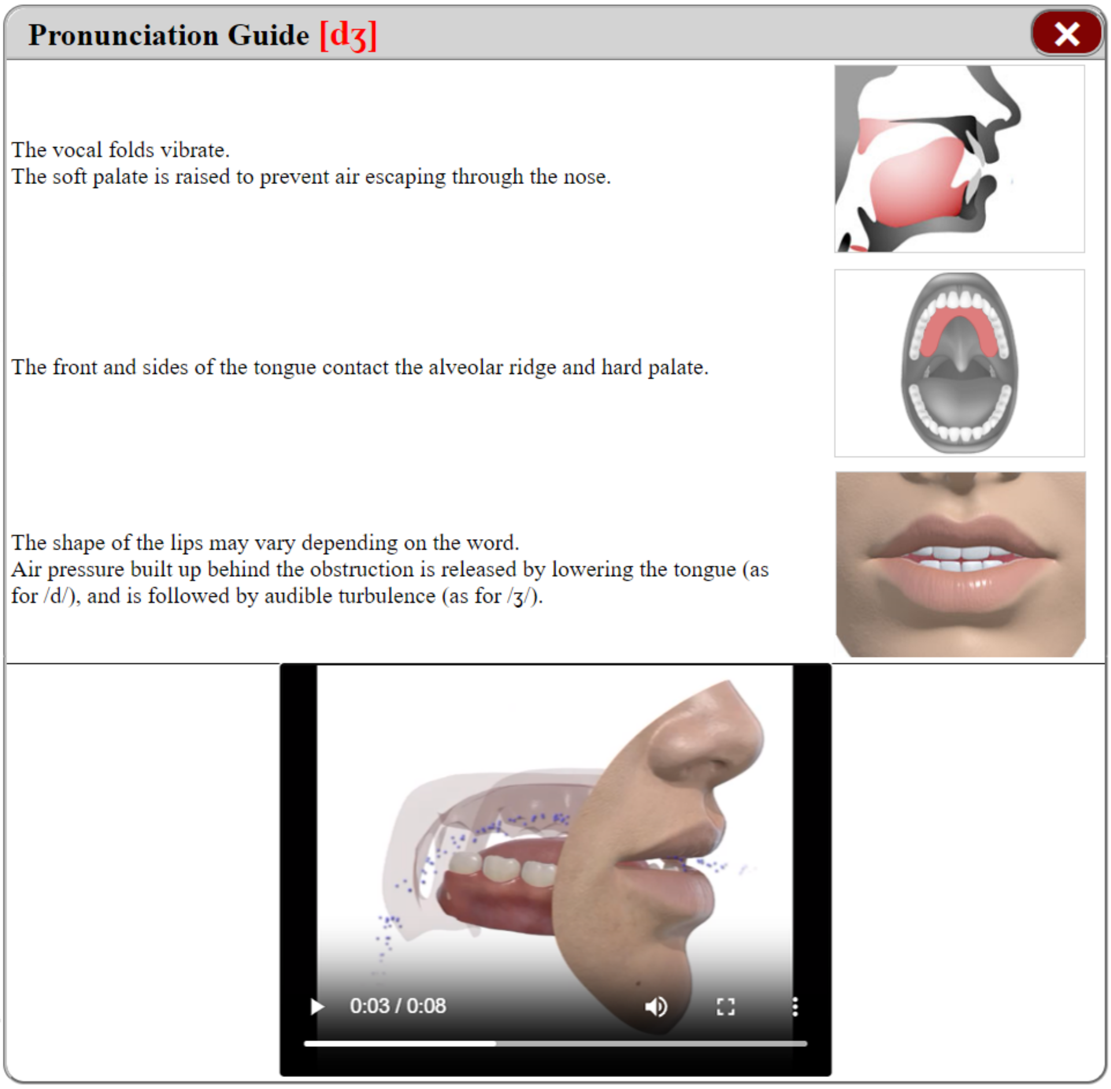
Accuracy Score	Fluency Score	Completeness Score	Pronunciation Score	Words Omitted
75	100	83	81	a, the

it100		was99			a	pleasure88						to100		be100		the	first90					to100		see94		the74		zoo57	
ɪt100		wɑz99			-	plɛ92			ʒər85			tu100		bi100		-	fɜrst90					tu100		si94		ði74		zu57	
ɪ	t	w	ɑ	z	-	p	l	ɛ	ʒ	ə	r	t	u	b	i	-	f	ɚ	r	s	t	t	u	s	i	ð	i	z	u
100	100	100	100	94	-	84	97	100	100	91	62	100	100	100	100	-	96	100	96	100	29	100	100	100	88	64	85	75	49

Note. Overall evaluation scores are presented in summarized form based on pronunciation accuracy, fluency, completeness, and overall pronunciation. Omitted and inserted words are also presented if applicable. Phonetical accuracy is graded at word, syllable, and phoneme levels and color-coded for ease of interpretation. Phoneme scores are hyperlinked to animated corrective feedback as shown in Figure 4.

Figure 4

Animated Pronunciation Instruction



Note. Animated pronunciation information is accessed by clicking on any phoneme in the evaluation scores shown in Figure 3. The instruction content can be customized to suite the target users. In this example, three paragraphs of instruction are presented with corresponding illustrations. A short video of the phoneme production can also be included.