https://doi.org/10.59865/abacj.2024.58

ALL-TALK: ENHANCING EFL PRONUNCIATION WITH MICROSOFT AZURE SPEECH SERVICES

Simon Moxon*

Abstract

This study introduces ALL-Talk, a web-based autonomous learning platform designed to enhance English-speaking skills among EFL students. Informed by extensive literature on second language speech influences, speaking anxiety, corrective feedback, and technology integration in language learning, ALL-Talk leverages Microsoft Azure's capabilities, including Text-to-Speech, Automatic Speech Recognition, Automatic Pronunciation Assessment, and immediate visual feedback mechanisms. ALL-Talk was evaluated over ten weeks with 17 EFL undergraduate students, focusing on enhancing their Business English communicative skills through improved fluency and pronunciation. Although changes in fluency between the pre-test and post-test were not statistically significant, $t(16) = 1.29$, $p = .215$, 95% CI [-2.19, 9.01], $d = 0.31$, male students improved significantly in overall pronunciation accuracy, $t(5) = 3.19$, $p = .024$, 95% CI [1.61, 15.06], $d = 1.30$. Additionally, both genders improved significantly in pronouncing /dʒ/, /z/, and /θ/. Preliminary evaluation and feedback indicate potential for ALL-Talk to support autonomous learning and speaking improvement in EFL contexts. However, future research should incorporate a longer evaluation period to yield more substantial research outcomes.

Keywords: Autonomous Learning, EFL Pronunciation, Speech Technology, Computer Aided Language Learning, Corrective Feedback

1. INTRODUCTION

Pronunciation and fluency are critical aspects of EFL learning, often assessed as key factors of language proficiency in standardized tests like IELTS. However, the mother tongue (L1) significantly influences second language (L2) speech (Bergmann et al., 2015), where L2 refers t*o any language learned after the acquisition of the first language. This influence often results in errors due to phonological differences between L1 and L2 (Gabriel, 2023; Ngo et al., 2023), such as substituting or excluding sounds absent in the L1 (Behr, 2022; Storkel, 2003; Van den Doel et al., 2018). For instance, Thais typically struggle with the English /θ/ sound, substituting it with /t/ (Sridhanyarat, 2017), and often weaken or exclude final consonant sounds (Isarankura, 2015). In contrast, Indonesians struggle with English characters representing multiple phonemes, as each consonant and vowel in Indonesian typically corresponds to a single phoneme (Leviakandella, 2022).

Despite their importance, pronunciation and fluency instruction are often marginalized in the classroom (Thomson & Derwing, 2015). This is partly due to instructors' limited knowledge or confidence in pronunciation pedagogy (Nushi & Sadeghi, 2021; Tavakoli &

* Dr. Simon Moxon is currently working as a lecturer in the School of Education & Liberal Arts at Walailak University, Thailand. He obtained a Ph.D. in Teaching Technology from Assumption University, Thailand. Email: simon.mo@mail.wu.ac.th

Hunter, 2018) and large class sizes (Alomari, 2024). Additionally, students' fear of making mistakes and facing public correction exacerbates speaking anxiety, further impeding language development (Ali & Fei, 2017; Bashori et al., 2022; Kusuma et al., 2022; Moxon, 2021; Wiboolyasarin et al., 2023).

In business contexts, pronunciation and fluency are essential for effective interaction as business professionals often engage in public speaking activities where English is the lingua franca (Louhiala-Salminen & Kankaanranta, 2011). Mispronunciation and disfluency can lead to misunderstandings and reduced credibility, hindering negotiation processes and career advancement (Kostromitina & Miao, 2024).

This research aimed to conceptualize, design, and develop ALL-Talk, an autonomous learning platform for EFL students to practice English-speaking skills. While studies have explored ways to improve pronunciation through audio analysis tools, Automatic Speech Recognition (ASR), Automatic Pronunciation Assessment (APA), and speech therapy software (Behr, 2022; Moxon, 2021, 2023), these applications often lack the integration needed for a unified learning system. ALL-Talk addresses this gap by integrating these technologies into a comprehensive, autonomous learning platform.

Informed by an extensive literature review of L2 speech influences, speaking anxiety, corrective feedback, and technology integration, ALL-Talk seeks to enhance L2 speech through a user-friendly integration of Text-To-Speech (TTS), ASR, APA, and visual feedback mechanisms. These mechanisms offer immediate corrective feedback to help students pinpoint pronunciation errors and visualize the stress and intonation patterns crucial for conveying confidence and authority in business settings (Kostromitina & Miao, 2024).

2. LITERATURE REVIEW

Among the four language skills, speaking is often prioritized by L2 learners (Leong & Ahmadi, 2017), with fluency and nativelike pronunciation as primary goals. The concept of fluency has attracted considerable debate concerning its definition, pedagogy, and assessment. Lennon (1990) distinguishes between two definitions: the "broad sense," encompassing overall language proficiency, and the "narrow sense," isolating fluency as a specific aspect of speech independent of vocabulary and grammar. This study adopts Lennon's "narrow sense" definition, focusing on fluency as a critical component in evaluating speaking proficiency. The literature review explores L1 influences, speaking anxiety, and the use of computer-aided language learning (CALL) systems to enhance L2 speaking proficiency.

2.1 The Influence of L1

Lai et al. (2009) identified five primary mechanisms through which L1 influences L2 pronunciation: (1) Fusion, characterized by the blending of L1 intonation and rhythm into L2 expression; (2) Absence, where specific L2 phonemes do not exist in the L1; (3) Substitution, whereby learners replace L2 sounds with acoustically similar L1 sounds; (4) Simplification or complexity, involving the addition or omission of consonants; and (5) Epenthesis, the insertion of consonants or vowels in order to comply with L1 phonotactic rules. L1 intonation and stress patterns can also significantly hinder L2 acquisition, leading to perceptions of dysfluency among listeners (De Jong, 2018). Therefore, identifying and correcting such L1 influences is crucial to prevent the fossilization of pronunciation errors (Hincks, 2003; Huang & Jia, 2016).

2.2 Speaking Anxiety

Classroom speaking anxiety and feedback provision are pivotal areas of concern in L2 acquisition. These phenomena have been explored through various theoretical lenses, highlighting the dynamics between learner discomfort and pedagogical strategies (Ali & Fei, 2017; Bashori et al., 2022; Hincks, 2003; Huang & Jia, 2016; Wiboolyasarin et al., 2023). Classroom speaking anxiety stems from the fear of speaking in front of peers, negative judgment of academic ability, or of making mistakes (Ali & Fei, 2017; Hincks, 2003), and is exacerbated by feedback and error correction, especially when conducted publicly or during oral presentations (Huang & Jia, 2016; Wiboolyasarin et al., 2023). Such anxiety deters active participation, crucial for developing speaking fluency.

2.3 The Role of Corrective Feedback

Despite recognizing the potential negative aspects of providing corrective feedback during oral production, Huang and Jia (2016) argue that careful correction of errors can prevent the fossilization of linguistic errors. Alutaybi and Alfares (2024) support this view, advocating for smaller class sizes to allow instructors to monitor students and provide feedback more effectively.

The effects of feedback extend beyond error correction, influencing student confidence, anxiety, motivation, and their ability to process and apply the feedback (Foote et al., 2016; Moxon, 2021; Nushi & Sadeghi, 2021; Olson & Offerman, 2021; Rogerson-Revell, 2021; Wiboolyasarin et al., 2023). For instance, specific, instructive feedback, rather than an overall score, enables students to pinpoint and correct their errors, facilitating more effective learning (Moxon, 2021).

Wiboolyasarin et al. (2023) examined the oral corrective feedback preferences of 288 students studying Thai in Asian universities. They found that while students typically did not feel shame, some reported feeling humiliated. Moreover, excessive corrective feedback given in front of peers made students feel restricted and anxious. The authors concluded that corrective feedback, while beneficial to the class, should be given privately.

2.4 Visual Feedback

Visual feedback, such as waveforms and pitch contours, has been shown to improve both the perception and production of intonation and pronunciation (Behr, 2022; Hincks, 2003; Olson & Offerman, 2021). Olson and Offerman (2021) found that the Visual Feedback Paradigm (VFP) significantly reduced Voice Onset Time (VOT) among L2 learners who compared visual representations of their speech against those of native speakers. Moxon (2023) also underscored the advantages of employing visual feedback, especially for comparative purposes.

2.5 Technology Integration

Integrating technology into educational contexts has revolutionized approaches to mitigating L1 influences on L2 pronunciation. Among these advancements, ASR stands out for its efficacy in enhancing pronunciation skills. Research highlights its pivotal role in increasing student engagement in L2 oral activities (Behr, 2022; Haggag, 2018) and its capacity to assist learners from diverse linguistic backgrounds in improving their pronunciation (Haggag, 2018; Moxon, 2021; Ngo et al., 2023; Sun, 2023).

Golonka et al. (2014) examined various technologies used in L2 learning, assessing their impact on pronunciation enhancement, language production, and learner motivation. They concluded that speech analysis software, notably PRAAT, enables students to practice and scrutinize their pronunciation autonomously. These findings challenge earlier assertions about

PRAAT's complexity and inaccessibility (Brett, 2004; Setter & Jenkins, 2005), suggesting instead that students find the software accessible and beneficial for self-analysis.

Supporting the use of PRAAT, Behr (2022) found that Thai EFL learners improved their pronunciation of the eight English diphthongs after training with the software. Behr's study involved Thai EFL undergraduates who recorded 80 test words three times before and after a month-long practice session involving 160 words. Over 60% of participants reported that PRAAT was user-friendly, while nearly 77% acknowledged that its visual representations of waveforms and spectrograms were instrumental in identifying sound movements. Behr's findings regarding the efficacy of PRAAT's visual representations align with those of Hincks (2003) and Olson (2014).

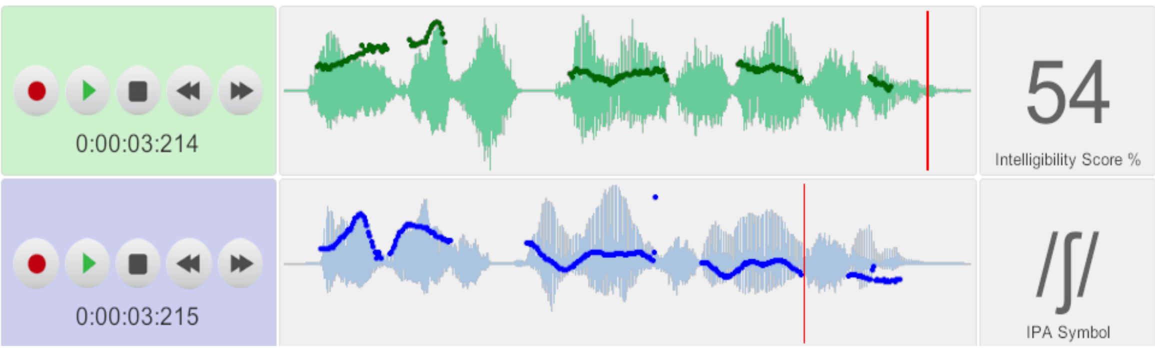
2.6 Features of Existing Platforms

The evaluation of current platforms was based on technical specifications, software reviews, and the author's personal observations. The primary platforms reviewed included BoldVoice, DuoLingo, ELSA, PRAAT, Pronounce, Pronunciation Coach 3D, Rosetta Stone, and SpeechAce.

2.6.1 Waveforms and Spectrograms

The visual representation of phonetic sounds through waveforms and spectrograms has repeatedly demonstrated its efficacy in enabling users to compare their speech visually with that of a native speaker (Behr, 2022; Olson, 2014). Platforms like Pronunciation Coach 3D, Rosetta Stone, and Say It English also foster some form of visual representation of speech (Moxon, 2023). However, Moxon (2021) highlighted a gap in the literature regarding the evidential superiority of visual representations over numerical data in correcting pronunciation errors. For example, as illustrated in Figure 1, Pronunciation Coach 3D provides a waveform of the target speech (lower waveform), displaying the relevant IPA symbol at specific points. Conversely, the user's speech (upper waveform) is evaluated with an overall intelligibility score, lacking phonetic-level feedback to compare against the target waveform. Differences in volume, intonation, or speech rate complicate pinpointing pronunciation errors through visual comparison alone, often necessitating a deeper understanding of phonology. Therefore, incorporating numerical feedback alongside visual representations could provide learners with more precise and actionable information, bridging the gap identified by Moxon (2021).

Figure 1 Pronunciation Coach 3D Waveforms



Note: Waveforms shown at near identical time markers (03:215) are not comparable due to dissimilar speech rates.

2.6.2 Corrective Feedback and Instruction

While language learning software increasingly incorporates ASR for pronunciation practice, the quality of corrective feedback and instructional content varies significantly across platforms. ELSA stands out by delivering instructional content for phoneme articulation, numerical pronunciation feedback, and corrective instruction for mispronounced phonemes (Nushi & Sadeghi, 2021). In contrast, Pronunciation Coach 3D, despite possessing similar features, lacks corrective instruction (Moxon, 2023). Meanwhile, SpeechAce offers minimal phonetic and corrective instruction (Moxon, 2021). Moreover, PRAAT provides no pronunciation instruction or corrective feedback.

Corrective instruction is crucial for L2 learners to remedy their errors independently (Huang & Jia, 2016; Ngo et al., 2023; Wiboolyasarin et al., 2023). Therefore, applications that offer both phonetic instruction and corrective feedback are likely more beneficial for autonomous learners.

2.6.3 Example Native Speech

A significant barrier for L2 learners is their limited exposure to the language within their learning or social environments (Wiboolyasarin et al., 2023). The evolution of TTS, with its enhanced natural speech capabilities, allows learners to hear nativelike speech from an array of AI-generated male and female voices, thereby engaging with the language in a more natural and accessible environment. However, uptake of this technology remains in its infancy, with few platforms offering a configurable TTS interface (Moxon, 2023). Customizable TTS settings, like the adjustment of speed, pitch, emotion, and accent, could enrich the learning experience by providing learners with auditory inputs matching their preferences.

2.6.4 Progress Tracking and Review

The capacity for learners to monitor and reflect on their performance is crucial for autonomous learning and self-improvement (Olson, 2014). However, existing platforms largely lack such features. Among the few exceptions, Pronounce stands out by offering a history of speaking tasks and error review progress, which can be shared with other users (Moxon, 2024). The ability for instructors to monitor student performance is a feature largely overlooked by other platforms.

The limited availability of monitoring features can be attributed to several factors. Many platforms are designed with the individual learner in mind, leading to a design architecture where review and evaluation data reside locally on the user's device, thus hampering access by educators or peers. Additionally, targeting specific operating systems like iOS, Android, and MS Windows, and relying on proprietary technology, exacerbates the challenge of achieving cross-device compatibility.

3. SYSTEM DESIGN

ALL-Talk was developed to provide autonomous L2 speaking practice with immediate feedback on speaking proficiency. Based on literature and platform analysis, six key features were identified for effective learning: 1) Delivery of target phrases using native-like speech (Moxon, 2023). 2) Visual representation of the target speech (Behr, 2022; Olson, 2014). 3) Visual representation of the user's speech for comparison with the target speech (Behr, 2022; Moxon, 2023; Olson, 2014). 4) Instant feedback (Nushi & Sadeghi, 2021). 5) Corrective instruction (Huang & Jia, 2016; Ngo et al., 2023). 6) Review and progress monitoring (Moxon, 2024; Olson, 2014).

ALL-Talk was developed using web technology and centralized data management to facilitate cross-platform compatibility. Central to the system's TTS, ASR, and APA aspects are the Azure Application Programming Interface (API) functions, providing natural speaking characters that generate speech from text (TTS) and speech analysis capabilities. These APIs deliver numerical feedback on native tone accuracy, fluency, completeness, and pronunciation accuracy at phrase, word, syllable, and phoneme levels. Azure's feedback mechanism, based on over 100,000 hours of native speech data (Eric-Urban, 2023), is designed to be instantaneous, consistent, and impartial.

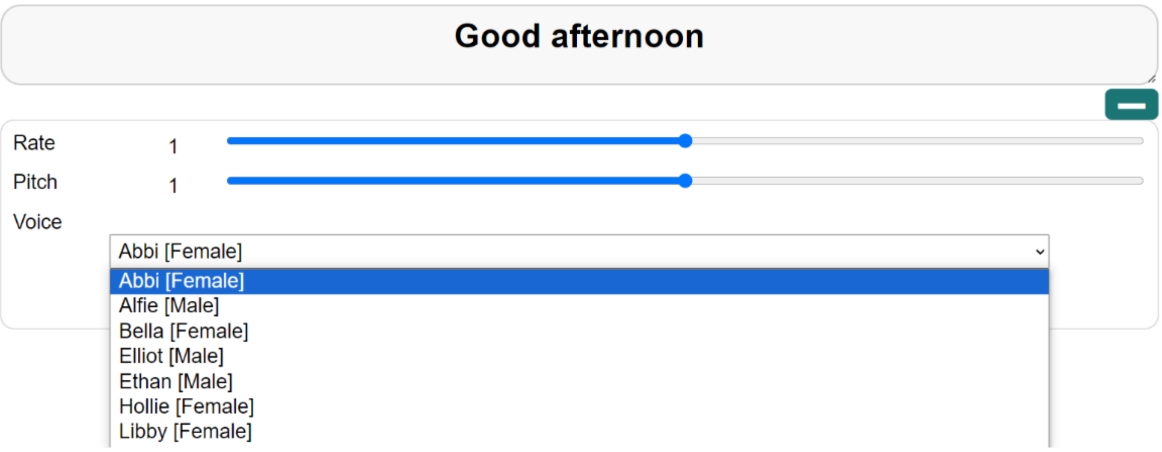
ALL-Talk adheres to relevant regulations, such as General Data Protection Regulations (GDPR), to ensure user privacy and data protection. These measures include Secure Sockets Layer (SSL) data transmission, encrypted storage of user authentication credentials, and obtaining informed verbal consent for data collection. The platform minimizes the storage of personal information, retaining only essential details, such as student ID, gender, and first name. Additionally, user activity, failed login attempts, and changes to authentication credentials are recorded and monitored regularly.

ALL-Talk is architecturally segmented into six core components: TTS, voice capture, audio-to-visual representation, speech evaluation and review, speaking task activity history, and user progress monitoring tools.

3.1 TTS

The literature identifies the lack of interaction with native speakers as a significant barrier to L2 speaking proficiency (Kusuma et al., 2022; Moxon, 2021). To address this, ALL-Talk integrates Azure's TTS functions, enabling users to hear target speech from a range of male and female native speakers. As Figure 2 illustrates, the interface allows the user to input a phrase and configure the speech based on the selected character, speech rate, and pitch. After processing by Azure, ALL-Talk converts the response to audio and displays it for playback with a corresponding waveform.

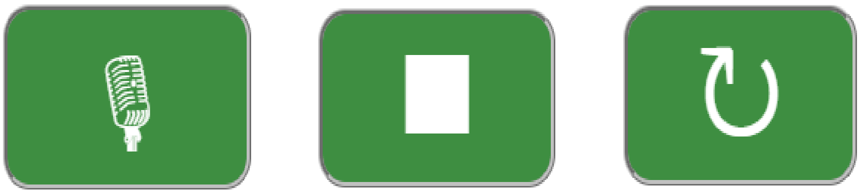
Figure 2 TTS Speech Character Options



3.2 Voice Capture

ALL-Talk is programmed to capture speech via the device's microphone. The user interface employs universally recognized symbols for starting, stopping, and reattempting recordings (Figure 3). Captured speech is converted to an audio file and visual waveform for playback and review (Figure 4).

Figure 3 Voice Capture Controls

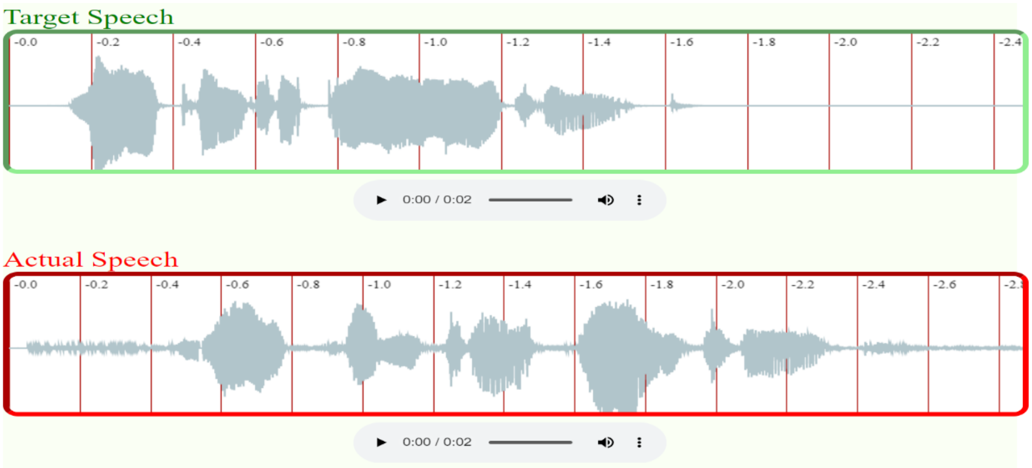


Note: Left to right: Start, stop, and reattempt buttons.

3.3 Audio to Visual Representation

In the reviewed system, two primary visual representations of audio emerged: waveform and spectrograph. The waveform representation is more widely adopted due to its ease of interpretation (Moxon, 2023). As Figure 4 illustrates, ALL-Talk transforms audio data into a graphical representation that plots time against sound pressure levels (pitch/amplitude), illustrating the waveform of the recorded audio and allowing the user to compare their speech visually with the target speech (Behr, 2022; Olson, 2014).

Figure 4 Waveform Diagrams and Audio Playback Controls



3.4 Speech Evaluation and Review

Paramount for autonomous learning is the ability to provide immediate feedback on performance (Pennington & Rogerson-Revell, 2018; Rogerson-Revel, 2021; Zou et al., 2023) so that strengths and weaknesses are easily identified (Moxon, 2021). Azure provides immediate speech evaluation, which ALL-Talk presents quantitatively across the linguistic units. To aid in interpreting evaluation scores, color-coded thresholds were defined: 0 to 39 (red, needing significant improvement), 40 to 59 (yellow, moderate proficiency), 60 to 79 (light-green, satisfactory performance), and 80 to 100 (dark-green, high proficiency). These thresholds were aligned with the university grading scales (A, B-C, D, and F) familiar to the students. A conservative range for the lower threshold was set to minimize loss of confidence and motivation in weaker students.

Additionally, as illustrated in Figure 5, omitted or inserted words are listed in the overall feedback, with omitted words highlighted in orange. Inserted words are those spoken out of order, including accidental or intentional repetition.

Figure 5 Speech Evaluation Feedback

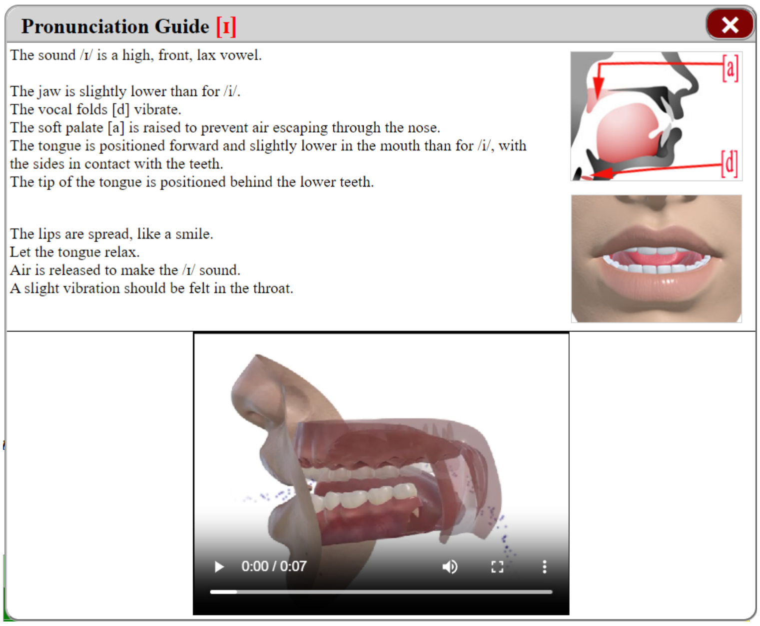
Accuracy Score	Fluency Score	Completeness Score	Pronunciation Score	Words Omitted
74	98	83	80	be

we^{100}		will^{100}			be	starting^{88}							from^{100}				scratch^{55}				
wi^{100}		wɪl^{100}			-	stɑr^{100}				tɪŋ^{8}			frʌm^{100}				skrætʃ^{55}				
w	i	w	ɪ	l	-	s	t	ɑ	r	t	ɪ	ŋ	f	r	ʌ	m	s	k	r	æ	tʃ
100	100	100	100	100	-	100	100	100	100	24	0	0	100	100	100	100	34	9	30	92	67

To enhance corrective feedback, ALL-Talk incorporates instructional content for each phoneme, which, as exemplified in Figure 6, includes images and videos to illustrate the movements of the mouth, tongue, teeth, and airflow. This instructional content can be configured to cater to learners' linguistic and age-specific needs. While this prototype version provides numerical feedback and instructional content for the production of each phoneme, it lacks features that guide the user through a structured remedial practice of problem sounds in isolation (Moxon, 2024; Nushi & Sadeghi, 2021). However, future enhancements could provide such functionality.

ALL-Talk offers three styles of evaluated speaking tasks, offering a comprehensive approach to pronunciation training. These consist of 1) user-entered short phrases, 2) instructor-set short phrases, and 3) audio blogs, which the instructor can set for students to discuss a topic or reflect on their weekly learning experiences. While all three tasks assess pronunciation accuracy, the audio blog emphasizes speaking fluency and reading accuracy skills.

Figure 6 Phoneme Production Guide

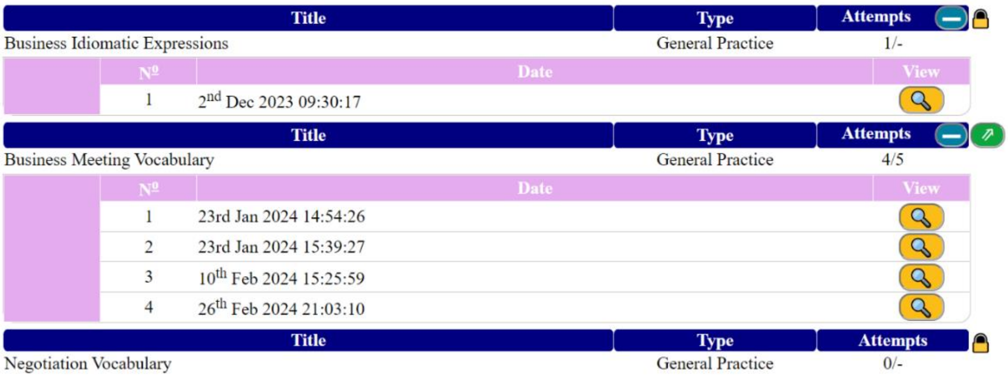


3.5 Speaking Task Activity History

Providing students with a means to review previous assessment attempts, scores, and feedback enhances their learning experience and performance (Moxon, 2024). It encourages self-regulated learning by allowing students to pinpoint improvement areas and actively monitor their progress (Olson, 2014).

For each speaking task, ALL-Talk stores the captured speech, transcript, and evaluation scores within the user's personal profile for subsequent review and reflection. This information is compiled into a personal progress history list, accessible exclusively by the individual, as depicted in Figure 7. Instructors can also monitor student performance via their progress monitoring tools, detailed in the section on Progress Monitoring Tools.

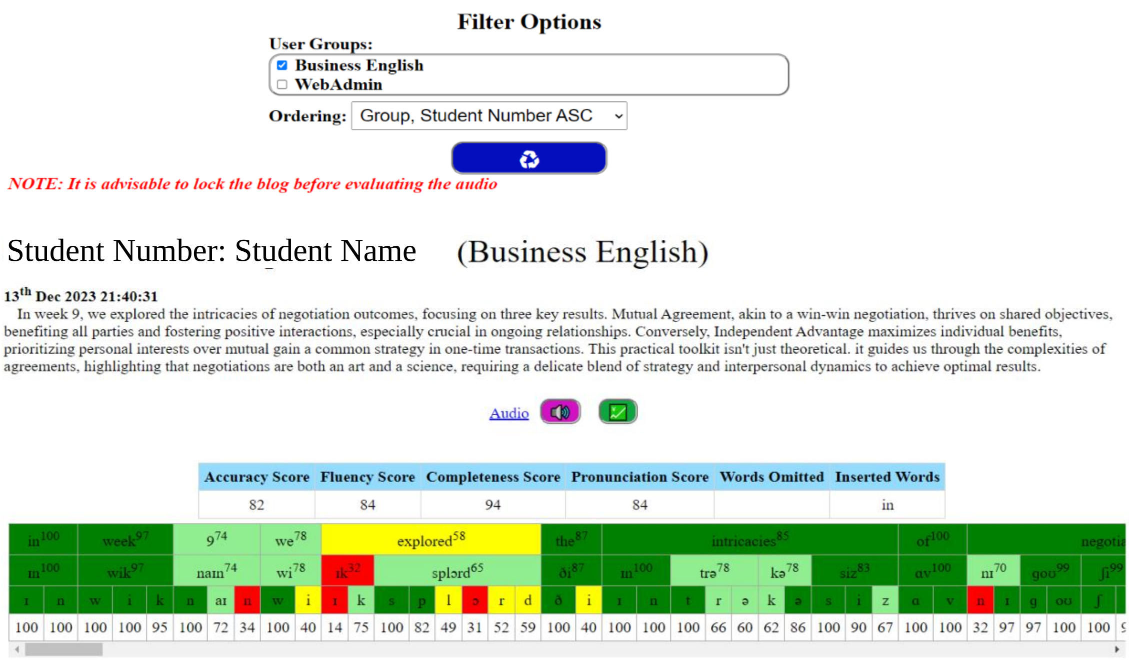
Figure 7 Speaking Task Progress History



3.6 Progress Monitoring Tools

ALL-Talk enables instructors to view all attempts for each speaking task within the relevant task editor. As illustrated in Figure 8, instructors can listen to student speech and

Figure 8 Speaking Task Attempts and Evaluations



review their evaluation scores. This feature allows instructors to quickly identify and address common or recurrent pronunciation errors while ensuring students record their speech correctly (Moxon, 2024).

For closer monitoring, ALL-Talk incorporates a reporting interface that facilitates instructors in aggregating performance data across multiple speaking tasks or specific phrases and phonemes. The reports provide comprehensive insights into students' fluency, accuracy, and pronunciation at various linguistic levels.

The report interface uses a two-stage process. Figure 9 depicts the first stage, in which the instructor selects the source of their report data (the speaking tasks). Figure 10 shows the second stage, where the instructor selects the output format of the report.

Figure 9 Progress Report Data Selection Criteria

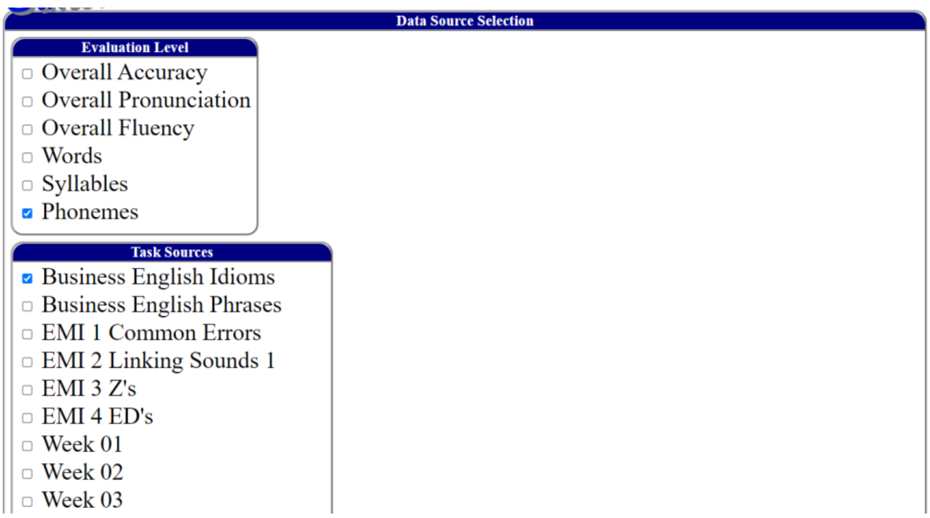
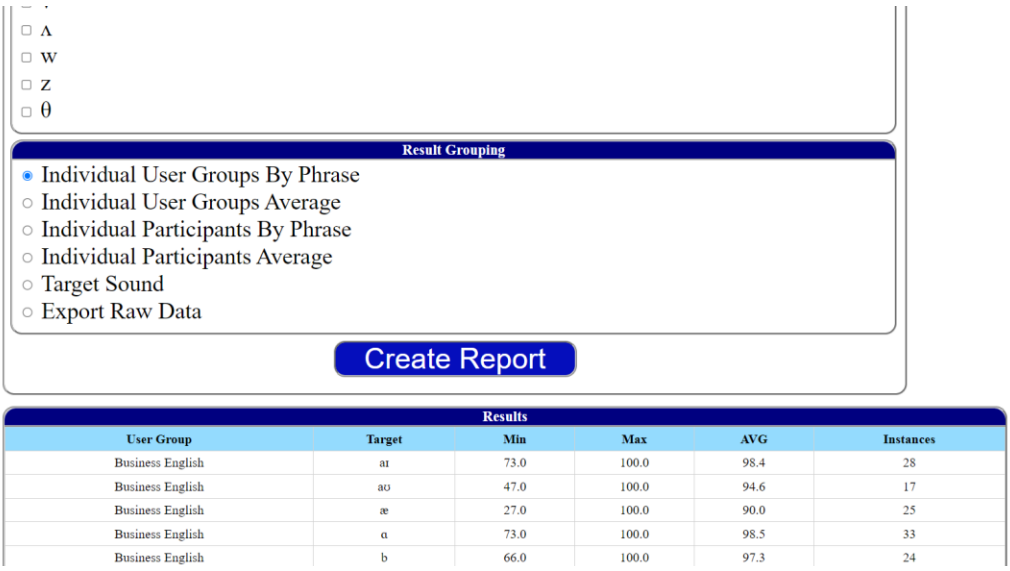


Figure 10 Progress Report Output Options



3.7 Validation and Future Enhancements

Given the prototype nature of ALL-Talk and the broad range of student devices used, a familiarization period was conducted with the target group to assess device compatibility before full implementation. Students could report issues during this period by relaying comments and screenshots electronically or in person. They were also encouraged to offer suggestions for enhancing user experience and platform functionality. Additionally, the author monitored system usage and checked speech recording quality to ensure proper use and functionality. This initial testing period led to several improvements to address device-specific issues and unhandled user input.

Based on observations and student feedback, future scalability enhancements include integrating AI alongside Azure functions to provide an interactive speaking companion. Combined with extended nationality TTS voices and emotions, this integration would enable students to practice real-time conversation in different varieties of English, a valuable skill for business communication in a globalized society. Additionally, the integration could offer students real-time advice on improving grammatical accuracy and diction while allowing them to seek clarification on learning shortfalls. Future versions should also include an interface for configuring evaluation score thresholds to reflect users' proficiency levels (Moxon, 2023) and an enhanced voice capture interface that allows users to pause and resume recording (Moxon, 2024).

3.8 Comparison Against Existing Platforms

Based on the existing platforms reviewed for this study, Table 1 presents a comparative analysis of their main features against those of ALL-Talk.

Table 1 Feature-Based Comparison Between ALL-Talk and Reviewed Platforms

Feature	ALL Talk	Speech Ace	ELSA	Rosetta Stone	Pronunciation Coach 3D	Pronounce	PRAAT	Bold Voice
ASR	☑	☑	☑	☑	☑	☑	☑	☑
APA	☑	☑	☑	☑	☑	☑	☐	☑
TTS	☑	☑	☐	☐	☑	☐	☐	☐
TTS Configuration	☑	☐	☐	☐	☑	☐	☐	☐
Visual Feedback	☑	☐	☐	☑	☑	☐	☑	☐
Attempt History	☑	☑	☑	☑	☐	☑	☐	☑
Corrective Instruction	☑	☐	☑	☐	☐	☐	☐	☑
Central Monitoring	☑	☐	☑	☐	☐	☐	☐	☐
Multi-Platform	☑	☑	☑	☑	☐	☑	☐	☑

4. SYSTEM EVALUATION

The system was evaluated by seventeen English major students (11 females and 6 males, aged 19-21) enrolled in the only Business English course offered within the Liberal Arts program. As an elective, this course attracts students specifically interested in advanced English vocabulary related to business, making it a distinctive opportunity for system evaluation. Given its elective nature and exclusive offering, the sample size was naturally constrained by the number of enrolled participants. Despite this limitation, classroom research offers valuable insights into the real-world applications of educational technology, which larger-scale studies may overlook (Plonsky & Oswald, 2014). As Plonsky and Oswald argue, statistical significance is highly influenced by sample size, and large samples will almost always approach significance. However, effect sizes provide a more reliable measure of practical significance, particularly in smaller samples, where trends and meaningful effects may not reach statistical significance due to power constraints (Lakens, 2013; Mackey & Gass, 2022).

Small classroom samples are a natural limitation in educational settings, yet the use of effect sizes (Cohen's *d* and η^2) helps to contextualize the practical impact of interventions like ALL-Talk, highlighting that meaningful changes in fluency and pronunciation can occur even when *p*-values are not below conventional significance thresholds (Norris & Ortega, 2000).

For the purpose of this evaluation, effect size thresholds are based on the benchmarks proposed by Plonsky and Oswald (2014) and defined as small ($d \approx 0.4$), medium ($d \approx 0.7$), and large ($d \approx 1$) for between-group comparison, and small ($d \approx 0.6$), medium ($d \approx 1$), and large ($d \approx 1.4$) for within-group comparison.

4.1 Statistical Analysis

To investigate the efficacy of ALL-Talk on speaking fluency and pronunciation accuracy, a series of statistical analyses were conducted. First, paired sample t-tests were used to compare pre-test to post-test scores for fluency and pronunciation of the overall sample and separately by gender, allowing for the evaluation of within-group changes over time. Next, independent sample t-tests were utilized to compare gender differences in fluency and pronunciation scores at both pre-test and post-test stages, aiming to identify any initial disparities or differential effects of the intervention between males and females. To control for baseline proficiency and isolate the effect of gender on improvement scores, an Analysis of Covariance (ANCOVA) with improvement scores as the dependent variable, gender as the fixed factor, and pre-test scores as the covariate, was employed. This methodological framework aligns with that of Fathi et al. (2020), who used similar statistical techniques to examine pre-test and post-test performance while controlling for initial proficiency levels.

Additionally, multiple regression analysis was conducted to explore how students' use of ALL-Talk features (e.g., log-ins, TTS, and speech evaluation) predicted improvements in pronunciation and fluency. Regression analysis is particularly suited for understanding how multiple predictors jointly influence a dependent variable and allows for greater insight into the relative importance of each predictor in educational technology research (Field, 2013; Tabachnick & Fidell, 2019).

Data were first examined for parametric testing suitability. Skewness and Kurtosis values for each variable were within acceptable ranges, and Q-Q plots confirmed that data points aligned with acceptable limits, indicating no violations of normality assumptions.

4.2 Analysis of Fluency Performance

Azure evaluates fluency scores based on the naturalness of the entire speech, considering speech rate and silence duration. This approach aligns with Lennon's (1990) "narrow sense" definition of fluency and the IELTS rubric indicators.

To assess the changes in speaking fluency, a paired sample t-test compared scores from week 1 and week 10. The difference between week 1 ($M = 76.41$, $SD = 13.97$) and week 10 ($M = 79.82$, $SD = 8.50$) was not statistically significant, $t(16) = 1.29$, $p = .215$, 95% CI [-2.19, 9.01]. The effect size, Cohen's $d = 0.31$, suggests that the observed improvement in fluency was small and should be interpreted with caution (Plonsky & Oswald, 2014).

Gender-specific analyses revealed no significant differences for males ($n = 6$) between week 1 ($M = 66.33$, $SD = 15.64$) and week 10 ($M = 74.50$, $SD = 10.00$), $t(5) = 1.61$, $p = .169$, 95% CI [-4.91, 21.25]. The effect size, Cohen's *d* = 0.66, suggests a small effect. For females (*n* = 11), no significant differences were found between week 1 (*M* = 81.91, *SD* = 9.78) and week 10 (*M* = 82.73, *SD* = 6.28), *t*(10) = 0.29, *p* = .782, 95% CI [-5.59, 7.23]. The effect size, Cohen's *d* = 0.09, indicated a negligible effect.

4.3 Gender Comparisons

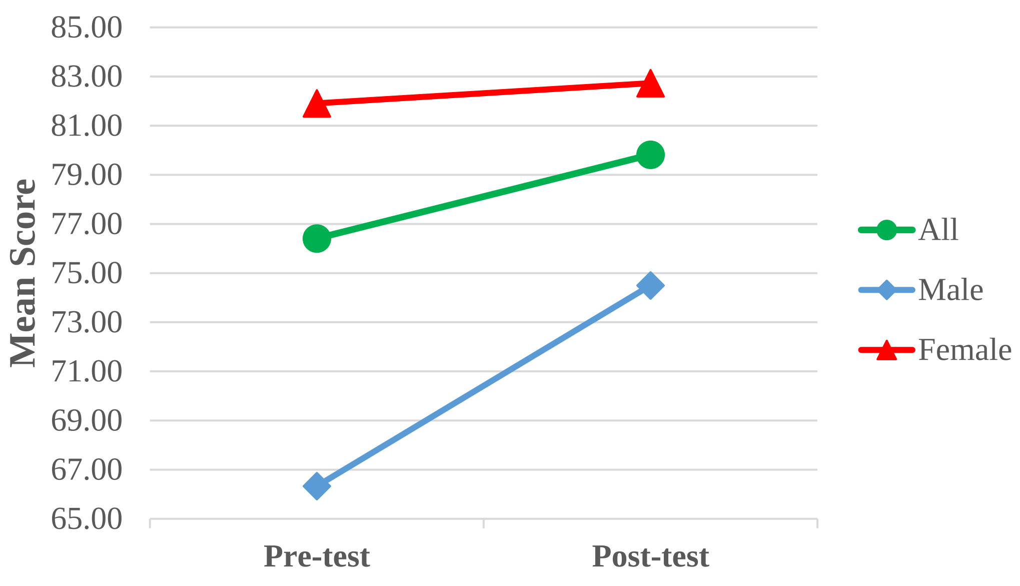
An independent sample t-test was conducted to compare mean fluency scores between genders at the pre-test and post-test stages. At the pre-test, males and females showed a significant difference in fluency scores, *t*(15) = 2.55, *p* = .022, 95% CI [2.53, 28.62], with females (*M* = 81.91, *SD* = 9.78) scoring higher than males (*M* = 66.33, *SD* = 15.64). The effect size was large, Cohen's *d* = 1.29, indicating a substantial difference in fluency between genders at the outset of the study.

At the post-test, females (*M* = 82.73, *SD* = 6.28) still scored higher than males (*M* = 74.50, *SD* = 10.00), although the difference in fluency scores between genders only approached statistical significance, *t*(15) = 2.10, *p* = .053, 95% CI [–0.12, 16.58]. The effect size remained large, Cohen's *d* = 1.07, suggesting that the magnitude of the difference was still considerable despite the lack of statistical significance (Plonsky & Oswald, 2014).

These results suggest that, while the difference in fluency scores between genders decreased in statistical significance over the course of the study, females consistently scored higher than males. The *p*-value reduction from pre-test to post-test suggests a trend toward narrowing the gender gap, potentially due to greater improvements among male participants. However, this trend was not statistically significant, and further research with larger samples is needed to confirm this observation.

As illustrated in Figure 11, overall improvements in fluency were observed in both genders.

Figure 11 Pre-test and Post-test Overall Fluency Scores



An Analysis of Covariance (ANCOVA) was conducted to evaluate the effect of gender on fluency improvement while controlling for any baseline differences in proficiency that could confound the relationship. The model included fluency improvement scores as the dependent variable, gender as the fixed factor, and pre-test scores as the covariate. Prior to the analysis, the assumptions of homogeneity of regression slopes and normality were tested and met, ensuring the validity of the ANCOVA results.

Descriptive statistics indicated that the mean improvement score was higher for males (M = 8.17, SD = 12.46) than females (M = 0.82, SD = 9.54). As Table 2 shows, the results revealed no significant main effect of gender on improvement scores after controlling for pre-test scores, $F(1, 14) = .60$, $p = .450$, $\eta^2 = .04$, suggesting that gender explained only a small portion of the variance in improvement scores. The covariate, pre-test score, was a significant predictor of improvement scores, $F(1, 14) = 21.09$, $p < .001$, $\eta^2 = .60$, indicating a large effect (Cohen, 1988), and suggesting that initial proficiency had a substantial impact on fluency improvement. The overall corrected model was significant, $F(2, 14) = 12.72$, $p = .001$, $\eta^2 = .65$, reflecting a large effect (Cohen), and explaining a significant portion of the variance in improvement scores.

These findings suggest that while gender did not have a significant effect on fluency improvement when controlling for baseline differences, initial proficiency levels were a strong predictor of progress. Specifically, for male students, the large effect size (Cohen's d = 0.66) suggests that the intervention had a meaningful impact on their fluency, though the non-significant results warrant cautious interpretation. Given the small sample sizes and non-significant findings, these conclusions should be interpreted tentatively, and future research with larger samples is necessary to validate these trends.

Table 2 ANCOVA Results for Fluency Improvement, Controlling for Pre-Test Scores

Source	*SS*	*df*	*MS*	*F*	*p*	Noncent Parameter	Observed Power [b]
Corrected Model	1223.17 [a]	2	611.56	12.72	.001	25.48	.986
Intercept	1171.78	1	1171.78	24.37	.000	24.38	.996
Pre-Test Score	1013.52	1	1013.52	21.09	.000	21.09	.989
Gender	29.01	1	29.01	.60	.450	.60	.112
Error	672.95	14	48.07				
Total	2094.00	17					
Corrected Total	1896.12	16					

a. $\eta^2 = .65$
b. Computed using alpha = .05

4.4 Analysis of Pronunciation Performance

Pronunciation scores, evaluated by Azure, reflect how closely each phoneme in the speech matches that of a native speaker. Syllable, word, and overall scores are aggregated from phoneme-level scores.

To assess the changes in pronunciation accuracy over time, a paired samples t-test was conducted comparing scores from week 1 and week 10. The results indicated no statistically significant difference between week 1 (M = 75.88, SD = 10.69) and week 10 (M = 82.00, SD = 10.26), $t(16) = 1.87$, $p = .080$, 95% CI[-.82, 13.05]. The effect size, Cohen's d = 0.45, suggested a small effect.

Gender-specific t-test scores revealed significant differences in pronunciation accuracy for males between week 1 (M = 71.83, SD = 10.87) and week 10 (M = 80.17, SD = 8.01), $t(5) = 3.19$, $p = .024$, 95% CI [1.61, 15.06]. The effect size, Cohen's d = 1.30, was large, indicating a substantial improvement in pronunciation accuracy for males (Plonsky & Oswald, 2014). However, this conclusion should be considered in light of the small sample size. No significant differences were found for females between week 1 (M = 78.09, SD = 10.43) and week 10 (M

= 83.00, *SD* = 11.54), *t*(10) = 1.00, *p* = .342, 95% CI [-6.05, 15.87]. The small effect size, Cohen's *d* = 0.30, suggests a minimal improvement in pronunciation accuracy for females.

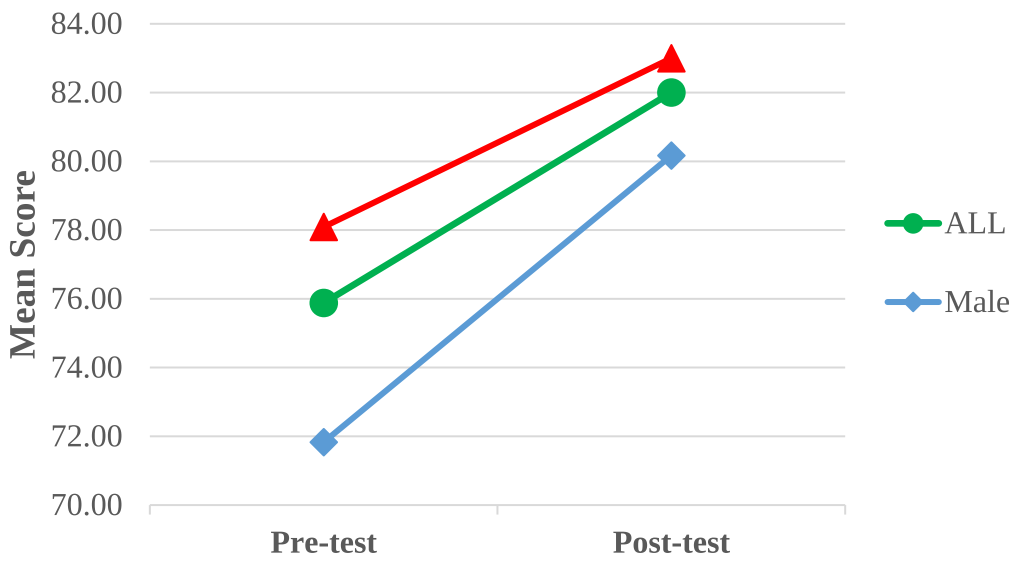
4.5 Gender Comparisons

An independent sample t-test was conducted to compare mean pronunciation scores between genders at the pre-test and post-test stages. At the pre-test, no significant difference in pronunciation scores was detected between females (*M* = 78.09, *SD* = 10.43) and males (*M* = 71.83, *SD* = 10.87), *t*(15) = 1.17, *p* = .262, 95% CI [–5.18, 17.70]. A medium effect size, Cohen's *d* = 0.59, was observed, suggesting that there was a moderate practical difference between the groups, even though the statistical significance was not achieved (Plonsky & Oswald, 2014).

For pronunciation scores at the post-test, there was no significant difference between genders, *t*(15) = 0.53, *p* = .603, 95% CI [–8.52, 14.19]. Females (*M* = 83.00, *SD* = 11.54) and males (*M* = 80.17, *SD* = 8.01) continued to have similar pronunciation accuracy. The effect size, Cohen's *d* = 0.27, was small.

As illustrated in Figure 12, although the improvement in pronunciation accuracy was not statistically significant for females, an overall improvement was observed in both genders.

Figure 12 Pre-test and Post-test Overall Pronunciation Scores



An Analysis of Covariance (ANCOVA) was conducted to evaluate the effect of gender on pronunciation improvement, controlling for initial proficiency levels. The model included the pronunciation improvement score as the dependent variable, gender as the fixed factor, and pre-test scores as the covariate. Assumptions of homogeneity of regression slopes and normality were met, ensuring the validity of the results.

Descriptive statistics indicated higher mean improvement scores for males (*M* = 8.33, *SD* = 6.41) than females (*M* = 4.91, *SD* = 16.31). As shown in Table 3, there was no significant main effect of gender on improvement scores, *F*(1, 14) = .12, *p* = .735, η^2 = .01, indicating that gender explained only a minor portion of the variance in improvement. However, the covariate, pre-test score, was a significant predictor, *F*(1, 14) = 10.76, p = .005, η^2 = .44, indicating that initial proficiency played a substantial role in pronunciation improvement. The corrected model was significant, *F*(2, 14) = 55.58, *p* = .017, η^2 = .44, explaining approximately 44% of variance in improvement scores.

Table 3 ANCOVA Results for Pronunciation Improvement, Controlling for Pre-Test Scores

Source	***SS***	***df***	***MS***	***F***	***p***	**Noncent Parameter**	**Observed Power** [b]
Corrected Model	1291.06 [a]	2	645.53	55.58	.017	11.15	.768
Intercept	1485.44	1	1484.44	12.83	.003	12.83	.914
Pre-Test Score	1245.54	1	1245.54	10.76	.005	10.76	.862
Gender	13.78	1	13.78	.12	.735	.12	.062
Error	1620.71	14	115.77				
Total	3548.00	17					
Corrected Total	2911.77	16					

a. $\eta^2 = .44$
b. Computed using alpha = .05

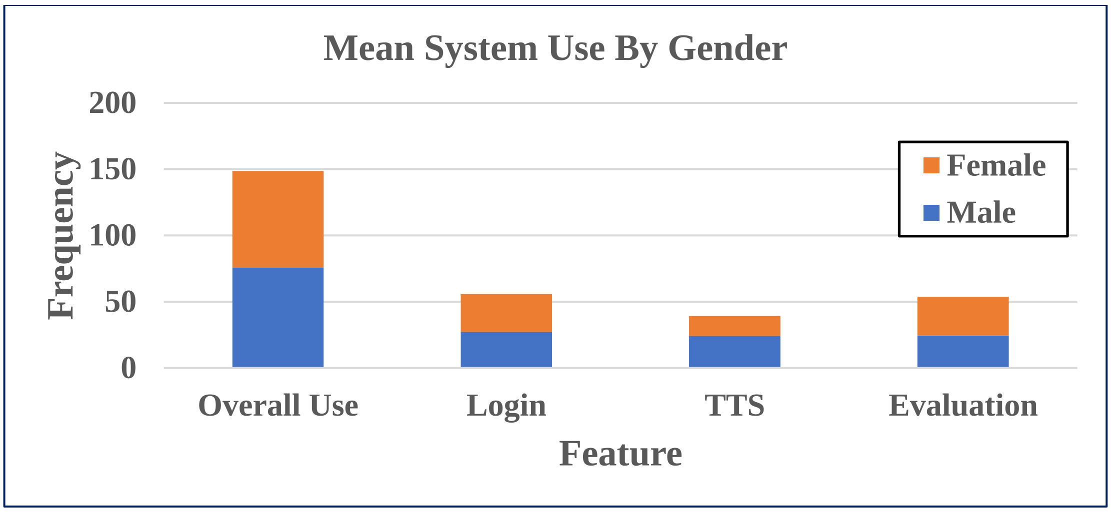
4.6 Analysis of Problem Phonemes

An evaluation of phoneme level pronunciation scores from the first audio blog task revealed three problematic phonetic sounds: /dʒ/, /z/, and /θ/, with mean scores below 75%. As Table 4 shows, a statistically significant improvement in these sounds was observed during the evaluation phase. The most notable improvement was in the pronunciation of the voiceless fricative consonant /th/ (/θ/), often mispronounced as /t/ by native Thai speakers (Sridhanyarat, 2017). Based on Plonsky and Oswald's (2014) benchmarks for L2 research, the effect sizes (Cohen's *d*) for /dʒ/, /z/, and /θ/ were 0.83 (medium to large), 0.55 (small to medium), and 0.72 (medium), respectively.

Table 4 Paired Sample t-Tests for Most Problematic Phoneme Sounds

Phoneme		*M*	*SD*	*SEM*	95% CI		*t*	*df*	*Sig. (2-tailed)*
					Lower	Upper			
dʒ	PostTest – PreTest	23.977	29.056	8.059	6.419	41.535	2.975	12	.012
z	PostTest – PreTest	8.675	15.835	3.959	.237	17.113	2.191	15	.045
θ	PostTest – PreTest	29.955	41.698	12.572	1.942	57.967	2.383	10	.038

Figure 13 Mean Use of ALL-Talk by Gender



4.7 Analysis of System Use

Descriptive statistics were used to evaluate ALL-Talk usage during the evaluation period. Figure 13 illustrates the overall usage and main feature usage categorized by gender. The results show that, while both genders used the system consistently, females were more active and evaluated their speech more frequently than males, utilizing the TTS and waveform features more extensively.

A multiple linear regression analysis examined the effect of three independent frequency-of-use variables (Log In, TTS, and Speech Evaluation) on pronunciation improvement. The model was significant, $F(3, 13) = 5.971$, $p = .009$, $R^2 = .579$, explaining approximately 58% of the variance in pronunciation improvement. As illustrated in Table 5, while Log In and Speech Evaluation were not significant predictors, the use of the TTS/Waveform was a significant predictor, $B = 0.681$, $SE = 0.250$, $\beta = 0.512$, $t(1) = 2.726$, $p = .017$. This suggests that the TTS and waveform features played a crucial role in enhancing pronunciation skills.

Table 5 Multiple Regression Analysis Prediction of Pronunciation Improvement Based on System Usage

Model	*B*	*SE B*	*β*	*t*	*p*	Collinearity Tolerance	VIF
Constant	-19.532	7.165		-2.726	.017		
Log In	.265	.220	.227	1.204	.250	.907	1.103
TTS/Waveform	.681	.250	.512	2.726	.017	.917	1.090
Evaluation	.210	.115	.338	1.826	.091	.943	1.060

A multiple linear regression analysis examined the effect of three independent frequency-of-use variables (Log in, TTS, and Speech Evaluation) on fluency improvement. The model accounted for approximately 37% of the variance in improvement scores ($R^2 = .370$), though this was not significant, $F(3, 13) = 2.54$, $p = .102$. As Table 6 illustrates, use of the speech evaluation feature was a significant predictor, B = 0.254, SE = 0.114, β = 0.508, t(1) = 2.239, p = .043, whereas Log In and TTS were not. This suggests that the speech evaluation feature played a crucial role in improving fluency. However, since the model did not explain a significant portion of the variance, other factors not included in the model may also be important.

Table 6 Multiple Regression Analysis Prediction of Fluency Improvement Based on System Usage

Model	*B*	*SE B*	*β*	*t*	*p*	Collinearity Tolerance	VIF
Constant	1.593	7.079		.225	.825		
Log In	-.348	.218	-.370	-1.599	.134	.907	1.103
TTS/Waveform	.252	.247	.235	1.022	.326	.917	1.090
Evaluation	.254	.114	.508	2.239	.043	.943	1.060

4.8 Student Feedback

At the end of the evaluation period, students completed a short questionnaire to express their opinions on using ALL-Talk for autonomous learning. The questionnaire consisted of seven questions evaluated on a five-point Likert scale. The distribution of student responses is presented in Table 7.

Table 7 Percentage Distribution of Responses to Likert Scale Questionnaire Items by Gender

Item	Gender	Strongly Agree (%)	Agree (%)	Neutral (%)	Disagree (%)	Strongly Disagree (%)
Using ALL-Talk was beneficial for practicing English speaking fluency and pronunciation outside of class.	Both	29.41	41.18	29.41	0.00	0.00
	M	50.00	33.33	16.67	0.00	0.00
	F	18.18	45.45	36.36	0.00	0.00
ALL-Talk provides a convenient way to practice speaking English.	Both	41.18	35.29	23.53	0.00	0.00
	M	66.67	16.67	16.67	0.00	0.00
	F	27.27	45.45	27.27	0.00	0.00
ALL-Talk is engaging and enjoyable to use.	Both	23.53	23.53	35.29	17.65	0.00
	M	33.33	16.67	50.00	0.00	0.00
	F	18.18	27.27	27.27	27.27	0.00
I feel motivated to use ALL-Talk for practicing English speaking.	Both	23.53	47.06	17.65	11.76	0.00
	M	33.33	50.00	16.67	0.00	0.00
	F	18.18	45.45	18.18	18.18	0.00
I feel less anxious when practicing English Speaking using ALL-Talk.	Both	17.65	52.94	17.65	11.76	0.00
	M	16.67	83.33	0.00	0.00	0.00
	F	18.18	36.36	27.27	18.18	0.00
I feel more confident in my pronunciation ability after using ALL-Talk.	Both	17.65	52.94	23.53	0.00	0.00
	M	16.67	66.67	16.67	0.00	0.00
	F	18.18	45.45	27.27	0.00	0.00
I feel more confident in my speaking fluency after using ALL-Talk.	Both	29.41	52.94	17.65	0.00	0.00
	M	50.00	33.33	16.67	0.00	0.00
	F	18.18	63.64	18.18	0.00	0.00

The feedback suggests that ALL-Talk was generally perceived as beneficial and convenient for practicing English speaking fluency and pronunciation, with a notable difference between genders. Overall, males expressed stronger agreement across most items, particularly regarding the convenience and benefit of using ALL-Talk outside of the classroom, with reduced anxiety cited as a key benefit.

5. DISCUSSION

The findings from this study underscore the practical significance of improvements in phoneme pronunciation for overall language proficiency. Accurate phoneme pronunciation is foundational for effective communication, and even marginal gains can significantly enhance a learner's intelligibility and confidence in speaking (Derwing & Munro, 2015; Levis, 2005). L2 pronunciation accuracy has been widely linked to communicative competence, particularly in professional settings such as business communication, where clarity and precision are essential (Derwing & Munro, 2015; Thomson & Derwing, 2015).

A notable outcome of this study is the gender disparity in pronunciation improvements, with males showing significant progress compared to their female counterparts. Males started with lower baseline scores (around 70%), which allowed for more noticeable improvements, whereas females, starting with higher baseline scores (around 80%), had less room for significant improvement. This phenomenon was explained by Baker-Smemoe et al. (2014) as a ceiling effect, where learners closer to proficiency show smaller incremental improvements.

Gender differences in learning styles and engagement with ALL-Talk may have also contributed to the results. Males utilized the TTS/Waveform features more extensively, which may have accelerated their phoneme acquisition through its detailed visual feedback on specific pronunciation targets. Greater interaction with these features among male participants aligns with previous research indicating that males often exhibit higher confidence and more positive attitudes toward using technology for learning compared to their female counterparts (Kahveci, 2010; Yau & Cheng, 2012). However, such differences can be moderated by factors such as nationality (Yu & Deng, 2022), and societal expectations (Bryla-Cruz, 2021).

Although, the study's small sample size (*n* = 17) limits the generalizability of the findings, this limitation is common in classroom-based research. Small samples reduce statistical power, making it harder to detect significant effects (Field, 2013), which underscores the necessity for caution when extrapolating these results to broader populations. This limitation highlights the importance of considering effect sizes alongside *p*-values, particularly in classroom-based research where larger samples may not be feasible.

Effect sizes help capture meaningful trends that could inform pedagogical practices, even when statistical significance is not achieved (Norris & Ortega, 2000). In this study, the medium effect size for males in fluency improvement (Cohen's $d = 0.66$) suggests that the intervention had a practical impact on this group, despite the lack of statistical significance, while the large effect size in pronunciation improvement for males (Cohen's $d = 1.30$) points to considerable gains from targeted pronunciation training (Mackey & Gass, 2022). These findings are consistent with Bashori et al. (2024) who found improved pronunciation through ASR-based training within a short five-week intervention period. Immediate phonetic feedback was attributed to the learning gains, mirroring the benefits observed in the current study's use of real-time pronunciation feedback tools.

In educational settings, especially those that focus on L2 acquisition, medium and large effect sizes highlight potential improvements that may not reach conventional significance due to sample size limitations. As highlighted by Plonsky and Oswald (2014), relying solely on *p*-values in small-sample studies can overlook practically important findings that might inform future pedagogical practices. This is particularly crucial in applied classroom research, where even modest improvements in learners' fluency and pronunciation can have meaningful implications for their communicative competence and confidence (Norris & Ortega, 2000).

The observed medium and large effect sizes in this study underscore that the ALL-Talk platform meaningfully contributed to students' pronunciation and fluency development. For example, the observed trends suggest that the TTS and waveform features were particularly beneficial for male students, facilitating improvement in both phoneme-level accuracy and fluency. Even though not all findings reached statistical significance, the practical gains observed can have a significant impact on classroom learning outcomes (Chapelle, 2003). These findings highlight the potential for technology-enhanced pronunciation tools to contribute effectively to language learning.

6. CONCLUSION

This research set out to design and develop ALL-Talk, a web-based platform aimed at enhancing English-speaking skills among EFL Business English students. Grounded in established learning theories, ALL-Talk's modules—including TTS, voice capture, audio-to-visual representation, speech evaluation and review—offered students tailored, technology-enhanced pronunciation and fluency practice. Over the ten-week evaluation, students demonstrated notable improvements in pronunciation and fluency, particularly with challenging phonemes such as /dʒ/, /z/, and /θ/, which are crucial for clear communication in professional settings.

However, the lack of significant improvement in fluency suggests that while phoneme pronunciation can be enhanced, this does not necessarily translate to enhanced fluency within the same evaluation period. Fluency may require extended practice and varied speaking contexts for notable progress. Longer-term use of ALL-Talk could lead to increased speaking confidence and more natural intonation patterns, critical for fluent and engaging speech.

The practical implications are substantial. Accurate pronunciation and fluency are linked to professional credibility and effective communication in business contexts. Mastering specific phonemes and developing fluent speech patterns can enhance overall intelligibility and confidence, better positioning students in global business environments. These skills are crucial for career advancement and professional success.

Despite promising outcomes and positive feedback, several limitations must be acknowledged. The small sample size (*n* = 17), limited participant demographics, and relatively short evaluation period restrict the findings' generalizability and applicability to broader EFL contexts. The use of effect sizes, which demonstrated practical significance despite the lack of statistical significance in some areas (e.g., fluency), provides valuable insights into how learners can benefit from technology-assisted language learning. Future research should involve larger, more diverse learner populations and extend the duration of interventions to better capture the long-term benefits of pronunciation and fluency practice.

ACKNOWLEDGMENTS

The author would like to thank the Human Research Ethics Committee of Walailak University for considering and approving the research used to evaluate ALL-Talk (Approval Number: WUEC-23-245-01).

REFERENCES

Ali, T., & Fei, W. (2017). Foreign language classroom anxiety among Iraqi students and its relation with gender and achievement. *International Journal of Applied Linguistics & English Literature, 6*(1), 305-310. https://doi.org/10.7575/aiac.ijalel.v.6n.1p.305

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

Alutaybi, M. M., & Alfares, N. S. (2024). EFL teachers' perspective of teaching large online classes: Issues and challenges. *Journal of Language Teaching and Research, 15*(1), 255-262. https://doi.org/10.17507/jltr.1501.28

Baker-Smemoe, W., Dewey, D. P., Brown, J., & Martinsen, R. A. (2014). Variables affecting L gains during study abroad. *Foreign Language Annals, 47*, 464-486. https://doi.org/10.1111/flan.12093

Bashori, M., Van Hout, R., Strik, H., & Cucchiarini, C. (2022). Web-based language learning and speaking anxiety. *Computer Assisted Language Learning, 35*(5-6), 1058-1089. https://doi.org/10.1080/09588221.2020.1770293

Bashori, M., Van Hout, R., Strik, H., & Cucchiarini, C. (2024). I can speak: Improving English pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching*. 1–19. https://doi.org/10.1080/17501229.2024.2315101

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. http://callej.org/journal/23-1/Behr2022.pdf

Bergmann, C, Sprenger, S. A., & Schmid, M. S (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychologica, 161*, 25-25. https://doi.org/10.1016/j.actpsy.2015.07.015

Brett, D. (2004). Computer generated feedback on vowel production by learners of English as a second language. *ReCALL, 16*(1), 102-113. 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

Chapelle, C. A. (2003). *English language learning and technology: Lectures on applied linguistics in the age of information and communication technology*. John Benjamins Publishing Company.

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

De Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly, 15*(3), 237-254. https://doi.org/10.1080/15434303.2018.1477780

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation Fundamentals: Evidence-Based Perspectives for L2 Teaching and Research.* John Benjamins Publishing.

Eric-Urban. (2023). *Transparency note and use cases for pronunciation assessment - azure AI services.* Microsoft. https://learn.microsoft.com/en-us/legal/cognitive-services/speech-service/pronunciation-assessment/transparency-note-pronunciation-assessment

Fathi, J., Derakhshan, A., & Torabi, S. (2020). The Effect of Listening Strategy Instruction on Second Language Listening Anxiety and Self-Efficacy of Iranian EFL Learners. *Sage Open*, *10*(2). https://doi.org/10.1177/2158244020933878

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

Foote, J. A., Trofimovich, P., Collins, L., & Urzúa Soler, F. (2016). Pronunciation teaching practices in communicative second language classes. *The Language Learning Journal*, *44*(2), 181-196. https://doi.org/10.1080/09571736.2013.784345

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

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*(1), 70-105. https://doi.org/10.1080/09588221.2012.700315

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

Hincks, R. (2003). Speech technologies for pronunciation feedback and evaluation. *ReCALL, 15*(1), 3-20. https://doi.org/10.1017/S0958344003000211

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

Kahveci, M. (2010). Students' perceptions to use technology for learning: Measurement integrity of the modified Fenneme-Sherman attitudes scales. *The Turkish Online Journal of Educational Technology, 9*(1). http://www.tojet.net/articles/v9i1/9121.pdf

Kostromitina, M. & Miao, Y. (2024). Listener perception of appropriateness of L1 and L2 refusals in English. *Studies in Second Language Learning and Teaching, 14*(2), 291-310. https://doi.org/10.14746/ssllt.38438

Kusuma, I. P. I., Pratiwi, N.P.A., Paramartha, A. A. G. Y., & Utami, I. A. M. I. (2022). Can e-portfolios reduce foreign language anxiety in online speaking courses during the Covid-19 pandemic? *Computer Assisted Language Learning Electronic Journal, 23*(2). 38-55. https://callej.org/index.php/journal/article/view/395/323

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

Lakens D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), 387-417. https://doi.org/10.1111/j.1467-1770.1990.tb00669.x

Leong, L., & Ahmadi, S. M. (2017). An analysis of factors influencing learners' English speaking skill. *International Journal of Research in English Education, 2*(1), 34-41. https://doi.org/10.18869/acadpub.ijree.2.1.34

Leviakandella, N. (2022). English students' problems in pronouncing unexist English vowels and consonants in Indonesian language. *Teaching English and Language Learning English Journal, 2*(1), 81-89. https://doi.org/10.36085/telle.v2i1.3328

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*(3), 369-377. https://doi.org/10.2307/3588485

Louhiala-Salminen, L & Kankaanranta, A. (2011). Professional communication in a global business context: The notion of global communicative competence. *IEEE Transactions on Professional Communication, 54*(3), 244-262. https://doi:10.1109/TPC.2011.2161844.

Mackey, A., & Gass, S. M. (2022). *Second language research: Methodology and design* (3rd ed.). Routledge.

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). *A review of the website Pronounce.com*. *Computer Assisted Language Learning Electronic Journal*, *25*(3), 195-212. https://callej.org/index.php/journal/article/view/22/397

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), 1-18. https://doi.org/10.1017/S0958344023000113

Norris, J. M., & Ortega, L. 2000. Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-Analysis. *Language Learning, 50*(3), 417-528. https://doi.org/10.1111/0023-8333.00136

Nushi, M., & Sadeghi, M. (2021). A critical review of ELSA: A pronunciation app. *Computer Assisted Language Learning Electronic Journal, 22*(3), 287-302. https://callej.org/index.php/journal/article/view/365/295

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

Pennington, M. C., & Rogerson-Revell, P. (2018). Using technology for pronunciation teaching, learning, and assessment. *Research and Practice in Applied Linguistics*, 235-286. https://doi.org/10.1057/978-1-137-47677-7_5

Plonsky, L. & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878-912. https://doi.org/10.1111/lang.12079

Rogerson-Revell, P. M. (2021). Computer-assisted pronunciation training (CAPT): Current issues and future directions. *RELC Journal, 52*(1), 189-205. https://doi.org/10.1177/0033688220977406

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

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

Storkel, H. L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research, 46*(6), 1312-1323. https://doi.org/10.1044/1092-4388(2003/102)

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*. https://doi.org/10.3389/fpsyg.2023.1210187

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Tavakoli, P., & Hunter, A.-M. (2018). Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research, 22*(3), 330-349. https://doi.org/10.1177/1362168817708462

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36*(3), 326-344. https://doi.org/10.1093/applin/amu076

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., Wiboolyasarin, 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

Yau, H. K., & Cheng, A. L. F. (2012). Gender difference of confidence in using technology for learning. *The Journal of Technology Studies, 38*(2), 79-79. https://doi.org/10.21061/jots.v38i2.a.2

Yu, Z., & Deng, X. (2022). A meta-analysis of gender differences in e-learners' self-efficacy, satisfaction, motivation, attitude, and performance across the world. *Frontiers in Psychology, 13*, 897327. https://doi.org/10.3389/fpsyg.2022.897327

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). https://doi.org/10.1177/21582440231193818