OPEN ACCESS

Middle East Research Journal of Humanities and Social Sciences
ISSN 2789-7761 (Print) | ISSN 2958-2040 (Online)
Frequency: Bi-Monthly
DOI: https://doi.org/10.36348/merjhss.2025.v05i03.002

KUWAIT SCHOLARS PUBLISHER
Website: http://www.kspublisher.com/
Email: office@kspublisher.com

Leveraging Google Voice Typing Technology to Enhance Pronunciation Skills Among English Language Majors at Hung Vuong University

Nguyễn Thành Long[1*], Nguyễn Thị Thanh Huyền[1], Đặng Thị Quỳnh Hoa[1]
[1]Hùng Vương University – Phú Thọ province

Abstract: Accurate pronunciation constitutes a pivotal component in enhancing communicative competence among learners of English as a foreign language. Nonetheless, a considerable number of students continue to encounter challenges due to suboptimal practice methods and the absence of immediate, constructive feedback. This study investigates the efficacy of Google Voice Typing—an automatic speech recognition (ASR) tool—in facilitating pronunciation improvement among English-major undergraduates. Employing a controlled experimental design, the research monitored participants' pronunciation accuracy through pre- and post-intervention assessments conducted over a four-week training period. The findings reveal substantial improvement, particularly in the articulation of monophthongs, diphthongs, triphthongs, and complex consonant sounds. Despite its benefits, the tool demonstrates limitations, including difficulties in accurately recognizing non-native accents and its inability to provide feedback on prosodic features such as stress and intonation. Accordingly, the study advocates for the integration of ASR technology with informed pedagogical strategies to optimize pronunciation instruction.

Keywords: Explicit Pronunciation Training, Automatic Speech Recognition (ASR), Google Voice Typing, English Phonetics, Immediate Feedback.

Research Paper

***Corresponding Author:**
Nguyễn Thành Long
Hùng Vương University – Phú Thọ province

How to cite this paper:
Nguyễn Thành Long *et al* (2025). Leveraging Google Voice Typing Technology to Enhance Pronunciation Skills Among English Language Majors at Hung Vuong University. *Middle East Res J. Humanities Soc. Sci*, *5*(3): 68-74.

Article History:
| Submit: 15.04.2025 |
| Accepted: 16.05.2025 |
| Published: 19.05.2025 |

1. INTRODUCTION

Precise pronunciation is a fundamental component in the acquisition of English as a second language (L2). Derwing and Munro (2005) emphasized that pronunciation is a critical determinant in the development of communicative competence and significantly impacts mutual intelligibility in interpersonal communication. Nevertheless, pronunciation instruction within traditional language classrooms often falls short of achieving optimal outcomes, particularly for learners who lack exposure to immersive, native-speaking environments. Against this backdrop, Automatic Speech Recognition (ASR) technology has emerged as a promising tool to support learners in refining their pronunciation skills. Among the most accessible and widely adopted ASR applications is Google Voice Typing, which allows learners to engage in pronunciation practice through interactive tasks that offer immediate, automated feedback. The primary objective of this study is to examine the effectiveness of Google Voice Typing in enhancing the pronunciation proficiency of English language majors at Hung Vuong University.

2. THEORETICAL FRAMEWORK

2.1. Explicit Pronunciation Training

Explicit Pronunciation Training (EPT) is a pedagogical methodology aimed at improving the pronunciation of English as a Foreign Language (EFL) learners through systematic instruction and focused, corrective feedback. Central to this approach is the notion that learners must develop metacognitive awareness of their phonetic errors and engage in intentional modification of their speech in order to enhance both articulatory precision and intelligibility in communicative contexts (Derwing & Munro, 2005).

Derwing and Munro (2005) contend that passive auditory exposure alone is insufficient for meaningful pronunciation improvement. Instead, learners require explicit, granular feedback addressing segmental features (such as individual phonemes), suprasegmental aspects (including stress, rhythm, and intonation), and articulatory patterns to align their speech more closely with target-language norms.

Expanding on this foundation, Skelton (2024) underscores the efficacy of EPT particularly for learners

whose native phonological systems diverge substantially from English. By enhancing phonological awareness and providing structured, data-informed feedback, EPT facilitates more accurate and intelligible pronunciation outcomes for these learners.

2.2. Automatic Speech Recognition (ASR) Technology in Language Education

Automatic Speech Recognition (ASR) refers to a sophisticated artificial intelligence technology that transcribes spoken input into written text and simultaneously evaluates pronunciation accuracy through advanced acoustic and linguistic modeling. ASR systems leverage an interdisciplinary framework combining digital signal processing, statistical phonetics, and machine learning to detect, interpret, and analyze learners' phonological patterns with considerable precision and scalability (González & Villanueva, 2019).

Within the domain of language education, ASR has garnered increasing attention as an innovative and impactful tool for facilitating pronunciation instruction, especially for learners of English as a Foreign Language (EFL). Its ability to provide instantaneous, automated feedback enables learners to identify and rectify phonetic deviations, thereby fostering heightened phonological awareness and articulatory accuracy. As a result, ASR contributes substantially to the advancement of oral fluency and overall communicative competence in the target language (McCrocklin, 2020).

2.3. The Application of ASR Technology in Pronunciation Training

The deployment of Automatic Speech Recognition (ASR) technology in pronunciation training has significantly expanded the horizons of foreign language pedagogy, particularly for learners of English as a Foreign Language (EFL). ASR facilitates instantaneous, individualized feedback, empowering learners to engage in self-directed and adaptive practice, even in the absence of direct instructor involvement (Skelton, 2024).

A key benefit of ASR lies in its ability to enhance learners' phonological awareness. González and Villanueva (2019) assert that ASR supports learners in autonomously tracking and rectifying their pronunciation, which is especially beneficial for addressing difficult or easily confusable phonemes influenced by L1 interference.

Moreover, ASR promotes the personalization of language learning. In contrast to traditional classroom settings, which are often constrained by time limitations and instructor-to-student ratios, ASR enables learners to practice at their own pace, without any restrictions on frequency or repetition. McCrocklin (2020) contends that this autonomy empowers learners to independently rectify errors that may otherwise elude detection in more conventional pedagogical environments.

However, ASR is not without its limitations. Neri et al. (2003) highlight that ASR systems frequently struggle to accurately process non-native speech, especially when phonological features diverge significantly from native speaker norms. Furthermore, current ASR technologies are unable to evaluate crucial suprasegmental features—such as intonation, stress, and rhythm—that are integral to achieving natural, fluent speech (Pennington & Rogerson-Revell, 2019).

Additionally, the efficacy of ASR feedback is influenced by various technical factors, including microphone quality, speaking rate, and environmental noise (Skelton, 2024). Consequently, to fully leverage the potential of ASR technology, its use should be complemented by instructional guidance and expert support.

In conclusion, ASR—particularly platforms such as Google Voice Typing—emerges as an invaluable tool for facilitating flexible and independent pronunciation improvement. Nonetheless, it should be regarded as a supplemental resource rather than a substitute for traditional pedagogical methods. The optimal integration of ASR with instructor-led guidance is likely to yield the most effective and authentic pronunciation outcomes for language learners.

3. HISTORY OF RESEARCH AND PREVIOUS STUDIES

3.1. The Evolution of Automatic Speech Recognition (ASR) Technology in Pronunciation Training

Automatic Speech Recognition (ASR) technology has witnessed substantial progress in the domain of pronunciation instruction. Early research, such as that conducted by Neri et al. (2003), concentrated on the fundamental capacity of ASR systems to recognize speech and provide automated feedback. At that stage, however, the technology was still in its formative years, and the systems were constrained by their inability to effectively process non-native accents.

With the advancement of artificial intelligence (AI) and machine learning, ASR systems have achieved considerable refinement, allowing for more accurate and nuanced pronunciation assessments, particularly when integrated with online learning environments (González & Villanueva, 2019). Despite these improvements, challenges persist in the accurate recognition of non-native speech due to inherent phonological and prosodic discrepancies between languages.

Furthermore, McCrocklin (2020) posits that interaction with ASR systems fosters a heightened metacognitive awareness of pronunciation, thereby enabling learners to engage in more effective self-correction and, ultimately, to refine their pronunciation with greater efficacy.

3.2. Previous Research on the Efficacy of ASR in Pronunciation Instruction

Empirical investigations have consistently affirmed the substantial effectiveness of Automatic Speech Recognition (ASR) technology in enhancing pronunciation proficiency among language learners. Neri *et al.*, (2003) demonstrated that learners employing ASR systems were able to identify and rectify phonetic errors with greater efficiency compared to those relying on traditional training methods.

In a similar vein, Pennington and Rogerson-Revell (2019) reported notable improvements in both phonemic accuracy and intelligibility among learners utilizing ASR technology. Nevertheless, they caution that, despite its advantages, ASR cannot yet fully replace human instructors, as it remains deficient in providing comprehensive feedback on suprasegmental elements, including stress patterns, intonation, and speech fluency.

Expanding on this notion, Skelton (2024) emphasized the synergistic potential of combining ASR with direct instructor guidance. He argued that learners who engage with ASR within a structured, pedagogically guided framework exhibit more rapid and substantial progress than those using ASR in isolation. This suggests that ASR should be viewed primarily as an auxiliary tool within goal-oriented instructional models rather than as an autonomous pedagogical solution.

3.3. Investigations on the Role of Google Voice Typing in Pronunciation Instruction

Google Voice Typing represents a prominent tool in the realm of Automatic Speech Recognition (ASR), enabling learners to input text via speech while receiving immediate, real-time feedback in the form of written text. This functionality allows learners to scrutinize and rectify pronunciation errors instantaneously. McCrocklin (2020) asserts that this tool facilitates a more visually intuitive error correction process compared to traditional pedagogical methods, proving particularly efficacious for challenging phonemes such as diphthongs, triphthongs, and consonants like /θ/, /ð/, /ʒ/, and /r/.

However, despite its merits, Google Voice Typing exhibits notable limitations, especially in processing non-native speech with high precision. The tool occasionally fails to detect subtle phonetic inaccuracies, which may undermine its utility for independent learning. While promising, the system is yet to achieve the level of nuanced accuracy required for comprehensive pronunciation enhancement across diverse learner profiles.

3.4. Synthesis of Findings from Prior Research

Upon synthesizing the findings from extant studies, several pivotal conclusions regarding the application of ASR in pronunciation instruction emerge:

ASR Facilitates Accelerated Pronunciation Error Detection and Correction: Research by Neri et al. (2003), McCrocklin (2020), and Pennington & Rogerson-Revell (2019) consistently underscores the capacity of ASR technology to enable learners to proactively identify and address pronunciation errors, significantly enhancing their phonetic accuracy over time.

Google Voice Typing is a Beneficial but Incomplete Tool: While Google Voice Typing provides instantaneous visual feedback, it does not adequately account for suprasegmental elements such as stress, intonation, and rhythm, which are crucial for achieving authentic, real-world communicative competence.

The Synergistic Integration of ASR and Instructor Guidance Yields Optimal Results: Studies by Skelton (2024) and González & Villanueva (2019) highlight that ASR should not be regarded as a standalone learning method. Instead, its integration within a well-structured curriculum, augmented by expert instructor guidance, leads to more significant and sustained improvements in pronunciation.

ASR Systems Require Refinement for Greater Efficacy with Non-Native Learners: A prominent challenge for ASR technologies lies in their capacity to accurately process non-native speech, which often deviates from native speaker norms due to differences in phonological and prosodic features. Scholars suggest that expanding training datasets and refining the underlying algorithms could substantially enhance the precision of ASR systems, particularly for non-native learners (González & Villanueva, 2019).

4. RESEARCH METHODOLOGY

4.1. Research Participants

This investigation was conducted with the participation of 40 undergraduate students enrolled in the English Language program at Hùng Vương University. These participants were in the second semester of their first academic year, possessing an intermediate level of English proficiency (A2-B1 according to the Common European Framework of Reference for Languages - CEFR). The selection of this particular cohort ensured that participants had a solid foundational understanding of phonetics, while still encountering challenges in achieving precise and fluid pronunciation.

The eligibility criteria for inclusion were as follows:

- The participants had no prior experience using Google Voice Typing for pronunciation practice.
- The participants did not exhibit any auditory impairments or significant speech disorders that could hinder their ability to pronounce words accurately.
- The participants demonstrated a willingness to engage in the study for a duration of four weeks

and adhere strictly to the prescribed practice schedule.

4.2. Research Design

The study employed a quasi-experimental research design to assess the impact of Google Voice Typing on the enhancement of students' pronunciation skills. The participants were randomly assigned to one of two groups:

- **Experimental Group**: Comprising 20 students who engaged in daily pronunciation practice utilizing Google Voice Typing.
- **Control Group**: Comprising 20 students who continued with conventional self-directed pronunciation practice without the benefit of ASR feedback.

Both groups underwent a pre-intervention pronunciation assessment (pre-test) at the onset of the study and a post-intervention assessment (post-test) following four weeks of practice. The assessment protocol consisted of the following components:

- **Single-word Reading Task**: Participants were asked to read a list of isolated words containing challenging phonemes commonly encountered by English learners as a foreign language (e.g., /θ/, /ð/, /ʒ/, /r/, diphthongs, triphthongs, etc.).
- **Passage Reading Task**: Participants were required to read a brief passage or dialogue, designed to evaluate both the fluency and accuracy of their pronunciation in context.

4.3. Data Analysis

The collected data were subjected to rigorous descriptive statistical analysis, with a primary focus on comparing the pre- and post-test scores of both the experimental and control groups. The accuracy of pronunciation was assessed using the precision of the phonetic feedback provided by the Google Voice Typing tool, which enabled real-time evaluation and correction of pronunciation errors during the practice sessions.

5. RESEARCH RESULTS

5.1. Improvement in Pronunciation Among Students

After four weeks of practice, the overall results indicated a notable improvement in pronunciation within the experimental group compared to the control group.

Table 1: Comparison of Pronunciation Improvement Between the Experimental and Control Groups

Evaluation Criteria	Before Practice (%)	After Practice (%)	Extent of Improvement (%)
Experimental Group			
Accurate Pronunciation of Individual Words Containing Difficult Phonemes (/θ/, /ð/, /ʒ/, /r/, Diphthongs, Triphthongs, etc.)	60%	90%	+30%
Improving Fluency and Accuracy in Reading Passages and Short Dialogues	40%	80%	+40%
Control Group			
Correct Pronunciation of Individual Words Containing Difficult Sounds (/θ/, /ð/, /ʒ/, /r/, Diphthongs, Triphthongs, etc.)	65%	70%	+5%
Improving Fluency and Accuracy in Reading Passages and Short Dialogues	30%	40%	+10%

The research findings reveal a significant improvement in students' pronunciation skills after four weeks of practice using Google Voice Typing. This enhancement was measured through pre-test and post-test evaluations, with criteria including vowel and consonant accuracy, word stress, and fluency in communication.

The extent of this improvement indicates that students not only refined their ability to pronounce individual sounds but also enhanced their pronunciation within the context of real-life communication.

5.2. Detailed Improvement by Phoneme Group (Experimental Group Only)

The pronunciation test included words and sentences containing single vowels, diphthongs, triphthongs, as well as challenging consonants that are commonly difficult for English as a Foreign Language (EFL) learners.

Table 2: Comparison of Detailed Improvement by Phoneme Group (Experimental Group Only)

Phoneme Group	Pre-Practice (%)	Post-Practice (%)	Improvement (%)
Single Vowels (/ɪ/, /æ/, ...)	60%	85%	+25%
Diphthongs (/aɪ/, /oʊ/, ...)	55%	87%	+32%
Triphthongs (/aɪə/, /ɔɪə/, ...)	50%	82%	+32%
Challenging Consonants (/θ/, /ð/, /ʒ/, ...)	58%	88%	+30%

5.3. Feedback from the Experimental Group on the Experience Using Google Voice Typing

The results of the student survey conducted at the end of the study show the following:

Feedback Content	Student Percentage (%)
The tool helps identify pronunciation errors	85%
Practice is more convenient due to instant feedback	80%
Increased confidence in English communication	75%
Incorrect recognition of some words when speaking quickly or with accent deviation	30%

- **85%** of students reported that Google Voice Typing helped them identify pronunciation errors they had previously overlooked.
- **80%** of students found the tool made practicing easier due to the instant feedback.
- **75%** of students noticed an improvement in their confidence when speaking English.
- **30%** of students mentioned that the tool sometimes failed to accurately recognize their pronunciation, particularly when speaking quickly or with a non-native accent.

5.4. Challenges in Using Google Voice Typing (Experimental Group Only)

In addition to the benefits, the study also highlights several challenges encountered when using this tool:

Accuracy of Speech Recognition

- Some students felt that the tool did not always accurately recognize difficult sounds, particularly when their intonation or stress patterns differed from native speakers.
- 30% of students reported that Google Voice Typing misidentified certain words, even when they believed their pronunciation was accurate.

Lack of Feedback on Stress and Intonation

- Google Voice Typing does not provide feedback on stress and intonation, which can affect real-life communication fluency.
  Technical Factors
- 25% of students encountered difficulties due to background noise, which impacted the accuracy of speech recognition.
- 20% of students indicated that an unstable microphone could lead to discrepancies in the practice results.

6. DISCUSSION

6.1. Efficacy of Google Voice Typing in Pronunciation Training

The findings of this study underscore the efficacy of Google Voice Typing as a robust tool in facilitating English language learners' pronunciation practice. One of the most pivotal advantages of this tool is its capacity to deliver immediate, real-time feedback on pronunciation accuracy. Throughout the practice sessions, students are able to instantly observe the recognition outcomes, enabling them to promptly identify and rectify any pronunciation discrepancies.

In addition, Google Voice Typing offers unparalleled flexibility, promoting a high degree of learner autonomy. Unlike conventional pronunciation teaching methods, which necessitate direct instructor guidance, this tool allows learners to engage in practice at their convenience, unhindered by time constraints or geographical limitations (McCrocklin, 2020). This degree of flexibility empowers students to tailor their practice schedules according to their personal needs, significantly increasing the frequency of pronunciation exercises—an indispensable factor in honing language proficiency (Pennington & Rogerson-Revell, 2019).

Furthermore, Google Voice Typing fosters the development of self-correction awareness among learners. Rather than relying solely on feedback from teachers or peers, students are encouraged to independently assess and refine their pronunciation errors during the practice process. This approach aligns with the principles of learner autonomy theory, which posits that learners' capacity for effective learning is enhanced when they are empowered to take ownership of their own educational journey (Skelton, 2024).

6.2. Challenges and Limitations

Despite the substantial advantages highlighted by the study, Google Voice Typing still presents several inherent limitations that deserve critical attention.

Primarily, the accuracy of the speech recognition system for non-native speakers remains a significant constraint. Since Google Voice Typing was developed based on native-speaker voice recognition models, it may encounter difficulties in accurately interpreting the pronunciation of learners whose speech patterns are influenced by their first language (L1 transfer) (González & Villanueva, 2019). The study's findings reveal that some participants reported instances where the tool failed to recognize words correctly, even though they believed their pronunciation to be accurate. This discrepancy can lead to an incorrect feedback effect, potentially eroding learners' confidence or causing them to misinterpret their pronunciation errors.

Moreover, Google Voice Typing does not offer comprehensive feedback on prosody, word stress, or fluency, focusing exclusively on the accuracy of individual phonemes. Intonation and word stress are crucial elements of spoken English, particularly in conversational contexts. However, the tool is not

equipped to assess or provide feedback on these aspects, which may limit its utility in preparing learners for authentic communicative situations (Pennington & Rogerson-Revell, 2019).

Another significant limitation is the influence of technical factors. The study found that 30% of participants encountered difficulties due to inconsistent microphone quality or background noise interference, which adversely affected the precision of speech recognition. This is consistent with Skelton's (2024) research, which underscores that the quality of the audio input can significantly impact the performance of automatic speech recognition systems.

6.3. Recommendations and Pedagogical Applications

Drawing from the advantages and limitations discussed, this study proposes several strategic recommendations designed to optimize the efficacy of Google Voice Typing in enhancing pronunciation skills for learners of English as a Foreign Language (EFL):

Augment Speech Recognition Accuracy for Non-Native Learners: It is imperative that developers broaden the speech corpus used to train the Automatic Speech Recognition (ASR) system, enabling it to more accurately accommodate a diverse array of accents and speech patterns, particularly those influenced by non-native prosody, intonation, and stress.

Integrate Comprehensive Feedback on Prosody and Intonation: Currently, Google Voice Typing predominantly offers feedback on word-level accuracy. Incorporating an advanced feature that evaluates sentence-level stress and intonation would not only augment learners' pronunciation precision but also foster more fluid and authentic communication in real-world interactions.

Synergize Google Voice Typing with Conventional Instructional Methods: Instead of treating the tool as an isolated learning method, it is recommended that educators integrate Google Voice Typing within the context of structured pronunciation lessons. This integration would facilitate the efficient use of the tool while also addressing critical aspects of pronunciation—such as stress patterns and intonation—that it currently cannot assess.

Promote Usage in Controlled Acoustic Environments: To ensure optimal recognition accuracy, learners should be encouraged to utilize Google Voice Typing in acoustically controlled, low-noise environments, equipped with high-fidelity microphones. This minimizes environmental interference and maximizes the system’s effectiveness in processing speech.

Develop a Systematic Pronunciation Training Program: Educators can design a graduated series of pronunciation exercises, structured from foundational to advanced levels. Such a program would enable students to progressively refine their skills, fostering both short-term improvements and long-term mastery, while allowing for comprehensive tracking of their development.

7. CONCLUSION

In conclusion, the findings of this study unequivocally highlight the efficacy of Google Voice Typing as an instrumental tool in facilitating the enhancement of English pronunciation, particularly in addressing the challenges posed by phonemes that non-native speakers commonly mispronounce. The average improvement rate ranging from 25% to 32% underscores the substantial impact of this tool in augmenting pronunciation accuracy. Moreover, Google Voice Typing provides instantaneous feedback, thus fostering greater learner autonomy, while affording students the flexibility to engage in pronunciation practice independently, without the necessity for continuous instructor supervision.

Nevertheless, the research also identifies certain limitations inherent to the tool. Specifically, its accuracy in recognizing non-native speech remains suboptimal, with notable gaps in feedback regarding prosody, stress patterns, and intonation—key elements crucial for effective communicative competence. Additionally, the tool's performance is susceptible to technical discrepancies, including issues related to microphone quality and environmental noise interference. As such, it is recommended that Google Voice Typing be employed as a supplementary resource rather than a standalone substitute for traditional pedagogical approaches.

This study contributes substantively to the burgeoning body of literature on the role of Automatic Speech Recognition (ASR) technologies in language acquisition, while simultaneously laying the groundwork for further explorations aimed at optimizing these technologies to enhance the quality of pronunciation instruction in educational settings.

REFERENCES

- Derwing, T. M., & Munro, M. J. (2005). *Second language accent and pronunciation teaching: A research-based approach*. TESOL Quarterly, 39(3), 379-397. https://doi.org/10.2307/3588485
- González, L. M., & Villanueva, M. A. (2019). Challenges in automatic speech recognition for second language learners: A review. *Computer-Assisted Language Learning*, 32(4), 315-330. https://doi.org/10.1080/09588221.2018.1523376
- McCrocklin, S. (2020). Dictation programs for second language pronunciation learning: Perceptions of the transcript, strategy use, and improvement. *Language Learning & Technology*, 24(2), 125-142. https://doi.org/10.14705/rpnet.2020.3.720

- Neri, A., Cucchiarini, C., & Strik, W. (2003). Automatic speech recognition for second language learning: Applications and challenges. *Speech Communication*, 41(1), 47-59. https://doi.org/10.1016/S0167-6393(02)00070-X
- Pennington, M. C., & Rogerson-Revell, P. (2019). *Using technology for pronunciation teaching, learning, and assessment*. In English pronunciation teaching and research. Research and Practice in Applied Linguistics. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-47677-7_5
- Skelton, L. (2024). *A Brief Guide to Speech Recognition Software in Tertiary Education*. Rere Āwhio – Journal of Applied Research & Practice, 4, 81-83. https://doi.org/10.34074/rere.00409