ELITE: English and Literature Journal
Volume 12 Number 1 June 2025 Page 68-79
Print ISSN: 2355-0821, Online ISSN: 2580-5215
Doi: 10.24252/elite. v12i1.56592

SPEECH-TO-TEXT TECHNOLOGY AS A TOOL FOR PRONUNCIATION ASSESSMENT AND SELF-CORRECTION IN EFL CONTEXTS: A PILOT STUDY

[1,2]James W Henry III
[1]Doshisha University, Kyoto, Jepang, [2]American College of Education, USA
akamatsucreative@gmail.com

ABSTRACT

This pilot study investigates the efficacy of speech-to-text (STT) technology as an objective feedback mechanism for pronunciation assessment and self-correction in English as a Foreign Language (EFL) classroom. The research involved 75 Japanese university students across three classes engaging with STT applications for two 25-minute sessions in a single semester. Data collected included successful transcription rates, specific pronunciation feature improvements, and student affective responses. Results indicate that STT software provides immediate, objective feedback that complements traditional pronunciation instruction methods, with students demonstrating measurable improvement in transcription success rates (from 38% to 62%) and reporting increased confidence in their speaking abilities. This pilot study contributes to the growing field of Computer-Assisted Language Learning (CALL) by offering a practical, accessible approach to pronunciation feedback that can be implemented with minimal technological requirements and time investment.
***Keywords**: Assessment, Computer-Assisted Language Learning (CALL), EFL instruction, English pronunciation, Speech recognition Technology*

ABSTRAK

Studi percontohan ini meneliti keefektifan teknologi speech-to-text (STT) sebagai mekanisme umpan balik objektif untuk penilaian pengucapan dan koreksi diri dalam kelas Bahasa Inggris sebagai Bahasa Asing (EFL). Penelitian ini melibatkan 75 mahasiswa universitas Jepang dari tiga kelas yang berinteraksi dengan aplikasi STT selama dua sesi 25 menit dalam satu semester. Data yang dikumpulkan mencakup tingkat keberhasilan transkripsi, peningkatan fitur pengucapan spesifik, dan respon afektif siswa. Hasil menunjukkan bahwa perangkat lunak STT memberikan umpan balik objektif dan langsung yang melengkapi metode pengajaran pengucapan tradisional, dengan siswa menunjukkan peningkatan terukur dalam tingkat keberhasilan transkripsi (dari 38% menjadi 62%) dan melaporkan peningkatan kepercayaan diri dalam kemampuan berbicara mereka. Studi percontohan ini berkontribusi pada bidang Computer-Assisted Language Learning (CALL) yang berkembang dengan menawarkan pendekatan praktis dan mudah diakses untuk umpan balik pengucapan yang dapat diimplementasikan dengan persyaratan teknologi dan investasi waktu minimal.
***Kata kunci**: , Computer-Assisted Language Learning (CALL, penilaian pengucapan bahasa InggriS, , umpan balik pengucapan, pengajaran EFL*

Received: 04 May 2025 **Revised: 15 May 2025** **Accepted: 27 May 2025**

How to cite: Henry, J. W. III. (2025). Speech-to-text technology as a tool for pronunciation assessment and self-correction in EFL contexts: A pilot study. *ELITE: English and Literature Journal,12(1), 68-79.*

INTRODUCTION

Language learning is a multifaceted process requiring the development of skills across listening, speaking, reading, and writing domains. Among these, pronunciation remains one of the most challenging aspects for English as a Foreign Language (EFL) learners, particularly for those whose first language phonological systems differ significantly from English. Traditional approaches to pronunciation instruction often rely heavily on subjective teacher feedback, which—while valuable may be inconsistent, time-constrained, or influenced by factors such as listener familiarity with accented speech.

The challenge of providing effective pronunciation feedback is further complicated in large classroom settings where individualized attention is limited. Additionally, learners may experience anxiety when required to practice pronunciation in front of peers, potentially hindering their willingness to engage in necessary practice. These challenges highlight the need for supplementary tools that can provide objective, immediate feedback and facilitate self-directed practice.

Recent technological advances have created new opportunities for pronunciation assessment and practice. Speech recognition technology, initially developed for practical applications such as dictation and voice commands, has evolved to a point where it can serve as a valuable tool in language education. This technology offers the potential for objective assessment based on intelligibility rather than accent reduction, aligning with contemporary communicative approaches to language teaching.

In the Japanese EFL context, pronunciation instruction faces challenges. The Japanese phonological system differs significantly from English in several key aspects, including the absence of certain phonemic distinctions such as /r/-/l/, limited consonant clusters, and different stress patterns. These structural differences, combined with traditional instruction methods that often emphasize grammar and reading over oral communication, create specific pronunciation challenges for Japanese learners of English that require targeted interventions.

Recent technological advances have dramatically expanded opportunities for pronunciation assessment and practice. Speech recognition technology has evolved significantly, with current research demonstrating its effectiveness across various EFL contexts (Liu et al., 2025; Sun, 2023). Recent studies have shown that automated speech recognition (ASR) systems can provide both corrective and confirmative feedback that supports pronunciation development while reducing learner anxiety (John et al., 2025; Xiao, 2025). This aligns with contemporary approaches that prioritize intelligibility and communicative confidence over native-like accent acquisition.

This pilot study examines how speech-to-text technology, specifically designed for transcribing spoken language into written text, can be repurposed as a pronunciation assessment and feedback tool in EFL contexts. The study was conducted with Japanese university students across three classes during two 25-minute sessions in a single semester, investigating both the practical implementation of this approach and its effects on student pronunciation accuracy and confidence. The small-scale, limited-resource approach of this study offers particular value for

instructors working in typical classroom environments with minimal technology access and significant time constraints.

The research addresses the following questions:

1. Can speech-to-text technology provide effective pronunciation feedback to EFL learners even in limited implementation contexts?
2. What specific pronunciation features can be effectively addressed using this approach?
3. How does the use of speech-to-text technology as a feedback mechanism affect learner attitudes toward pronunciation practice?

This pilot study contributes to the growing field of Computer-Assisted Language Learning (CALL) by examining the pedagogical applications of widely available technology, providing practical guidance for implementation, and assessing outcomes across multiple dimensions of language learning within a realistic classroom timeframe.

LITERATURE REVIEW

1. The Challenge of Pronunciation Instruction in EFL

Pronunciation instruction has historically occupied a contested position within language teaching methodologies. From the intense focus on native-like accuracy in the Audio-Lingual Method to the reduced emphasis during the Communicative Language Teaching movement, approaches to pronunciation have varied significantly (Celce-Murcia et al., 2010). Current research generally supports a balanced approach that emphasizes intelligibility and communicative effectiveness rather than accent elimination (Levis, 2018).

Munro and Derwing (2015) distinguish between intelligibility (the extent to which a listener can understand an utterance), comprehensibility (the perceived difficulty in understanding an utterance), and accentedness (how different a pattern of speech sounds from the local variety). This distinction is crucial, as research indicates that accentedness does not necessarily impede intelligibility or comprehensibility. The goal of pronunciation instruction, therefore, shifts from accent reduction to ensuring learners can be readily understood.

For Japanese learners of English specifically, several pronunciation challenges have been well-documented. These include the difficulty distinguishing /r/ and /l/ phonemes (Aoyama et al., 2004), challenges with consonant clusters that do not occur in Japanese phonotactics (Bradlow et al., 2001), and issues with English stress patterns due to the mora-timed nature of Japanese compared to the stress-timed nature of English (Ueyama, 2000). These specific challenges make Japanese EFL learners an appropriate population for testing pronunciation feedback mechanisms.

2. Feedback in Pronunciation Learning

Effective feedback is essential for pronunciation development. Lyster and Saito (2010) identify several feedback types in language learning contexts, including explicit correction, recasts, clarification requests, and metalinguistic feedback. Research suggests that while all

feedback types may be beneficial, their effectiveness varies depending on learner characteristics and the specific pronunciation features being addressed.

The timing of feedback also impacts its effectiveness. Immediate feedback allows learners to make connections between their production and the target form while the pronunciation attempt is still in working memory (Kartchava & Ammar, 2014). This immediate connection is difficult to achieve in traditional classroom settings but can be facilitated through technological solutions.

Baker and Burri (2016) note that affective factors significantly influence pronunciation learning, with anxiety potentially inhibiting willingness to practice and experiment with unfamiliar sounds. Technologies that provide private practice opportunities may help reduce these affective barriers, potentially increasing engagement with pronunciation practice.

3. Computer-Assisted Pronunciation Training (CAPT)

CAPT represents a specialized branch of Computer-Assisted Language Learning focused specifically on pronunciation development. O'Brien et al. (2018) note that CAPT systems typically incorporate speech recognition technology, acoustic analysis, and visual feedback to help learners identify and correct pronunciation errors. Advantages of CAPT systems include increased practice opportunities, reduced anxiety compared to classroom practice, immediate feedback, and the ability to focus on individual learner needs (Neri et al., 2008). However, challenges exist in terms of technology limitations, individual variation in speech patterns, and the need for pedagogical frameworks to guide implementation.

Most research on CAPT has focused on dedicated language learning software or specialized laboratory settings (Levis, 2007). Less attention has been paid to the potential of repurposing general-purpose speech recognition technology for classroom pronunciation practice, particularly in contexts with limited technological resources or instructional time constraints.

4. Automatic Speech Recognition in Language Learning

Automatic Speech Recognition (ASR) technology has evolved significantly over the past decade. Initially designed for practical applications like dictation and voice commands, ASR has increasingly been adapted for educational purposes (McCrocklin, 2019). Unlike systems designed specifically for language learning, general-purpose ASR software is typically "speaker-independent," meaning it attempts to recognize speech from any user without prior training.

Recent systematic reviews of ASR technology in EFL contexts have revealed significant advancements in both accuracy and pedagogical application (Liu et al., 2025). Unlike early speech recognition systems that struggled with accented speech, current ASR technologies demonstrate improved capability in recognizing and evaluating non-native pronunciation patterns, particularly for intermediate-level learners (Inceoglu et al., 2023). This technological evolution has expanded the potential applications of general-purpose speech recognition tools in EFL classrooms, even in settings with limited technological resources.This speaker-independence makes general-purpose ASR particularly suitable as a pronunciation assessment tool. Rather than accommodating learner

pronunciation patterns, these systems require the learner to adjust their pronunciation to match target language norms for successful recognition, providing an objective measure of intelligibility (Mroz, 2018).

Beyond mere pronunciation assessment, recent research has explored integrating ASR with peer feedback mechanisms to create comprehensive pronunciation learning environments (Evers & Chen, 2022; Rogti, 2025). These combined approaches leverage both the objective assessment capabilities of technology and the social learning benefits of peer interaction. Sun (2023) demonstrated that this integrated approach not only improved segmental accuracy but also enhanced overall communicative confidence among EFL learners. These findings suggest that the pedagogical value of speech recognition technology extends beyond error identification to include positive reinforcement of successful communication attempts.

Emerging research on the use of general-purpose ASR for pronunciation practice has shown promising results. Studies by McCrocklin (2016) and Wallace (2016) found that learners using ASR for pronunciation practice demonstrated improvement in specific phonological features and reported positive attitudes toward the technology. However, these studies also noted limitations in terms of ASR accuracy for heavily accented speech and the need for teacher guidance in interpreting ASR feedback.

5. Short-Term Interventions in Pronunciation Teaching

While longitudinal studies offer valuable insights into pronunciation development, shorter interventions can also yield meaningful results. Thomson and Derwing (2015) reviewed 75 pronunciation studies and found that even short-term interventions of a few weeks or less could produce measurable improvements, particularly when focused on specific pronunciation features rather than general accent reduction.

Lee et al. (2015) demonstrated that focused pronunciation instruction delivered in just four 50-minute sessions resulted in significant improvement in targeted features. Similarly, Sardegna (2011) found that strategy instruction in a single session could lead to improved pronunciation when paired with subsequent self-regulated practice.These findings suggest that even time-limited interventions, such as the two sessions employed in the current study, can potentially yield meaningful results when implemented effectively. This is particularly relevant for educational contexts where curricular constraints limit the time available for explicit pronunciation instruction.

RESEARCH METHODS

Research Design

This study employed a mixed-methods approach combining quantitative assessment of pronunciation accuracy with qualitative analysis of student experiences and perceptions. The research was conducted as a pilot study over a single semester, with data collection occurring for two 25-minute sessions in each participating class.

This mixed-methods approach follows recent methodological trends in pronunciation research that emphasize combining quantitative measurements of accuracy with qualitative assessment of learner experience (Sun, 2023). While comprehensive longitudinal studies provide valuable insights into pronunciation development, shorter interventions have also demonstrated measurable impacts on both segmental accuracy and learner confidence (Rogti, 2025). The present pilot study contributes to this emerging research paradigm by examining the efficacy of limited technological intervention in a typical classroom environment.

Participants

The participants included 75 Japanese university students enrolled in English oral communication courses at a private university in Kyoto, Japan. The students were distributed across three classes (approximately 25 students per class) and represented a variety of academic majors. All participants had studied English for at least six years through the Japanese secondary education system but reported limited opportunities for pronunciation-focused instruction or authentic speaking practice prior to university.

Materials and Technology

The primary technology used in this study was Nuance's Dragon Dictation application installed on Apple iPad devices. This speech-to-text application was selected based on several criteria:

1) Speaker-independence: The software does not adapt to user pronunciation, making it suitable for providing objective feedback.
2) Real-time transcription: The application displays text immediately after speech input, facilitating immediate feedback.
3) Accessibility: The free mobile version is easily accessible for both classroom and independent student use.
4) Usability: The interface is intuitive and requires minimal training to operate.

The selection of general-purpose speech recognition technology rather than specialized language learning software reflects recent research demonstrating the efficacy of commercially available applications in educational contexts (Xiao, 2025). This approach aligns with findings that suggest widely accessible technologies can reduce implementation barriers while providing comparable benefits to purpose-built systems (Inceoglu et al., 2023). Furthermore, the game-like structure of the activity draws on research showing that competitive elements can enhance learner engagement with pronunciation practice (Evers & Chen, 2022).

Additional materials included:

- A corpus of 15 practice sentences designed to target specific pronunciation features challenging for Japanese learners of English
- Brief questionnaires to gather qualitative data on student experiences

- One iPad per class for the activity implementation

Data Collection Procedures

The study implemented a two-phase data collection process:

- **Session 1 (Week 3 of semester).** Brief introduction to the technology and activity format (5 minutes). Students took turns attempting to have their spoken sentences accurately transcribed by the software (15 minutes). Teacher provided guidance on interpreting feedback and suggested strategies for improvement. Brief questionnaire on student reactions to the activity (5 minutes).
- **Session 2 (Week 12 of semester).** Review of the activity format (2 minutes). Students took turns attempting to have their spoken sentences accurately transcribed by the software (18 minutes). Final questionnaire on student experiences and perceived progress (5 minutes).

Due to time constraints and the "game" format of the activity, not all students attempted the same sentences, and the number of attempts varied across students. On average, each student made 3-5 attempts across the two sessions, with performance recorded for qualitative and quantitative analysis. The activity was structured as a "game" with students earning bonus points for successful transcriptions to increase engagement and reduce anxiety. Students who were waiting for their turn observed their peers' attempts, creating an environment of collaborative learning through observation. After each attempt, the instructor provided brief feedback on specific pronunciation features that may have contributed to transcription success or failure.

Data Analysis Methods

Quantitative data analysis included:

- Calculation of successful transcription rates (percentage of utterances correctly transcribed by the software)
- Comparison of success rates between Session 1 and Session 2
- Analysis of success rates for utterances containing specific challenging phonological features (e.g., /r/-/l/ distinctions)

Qualitative data analysis included:

- Thematic analysis of student questionnaire responses
- Documentation of student reactions during the activity
- Identification of patterns in pronunciation challenges and successful correction strategies

Given the exploratory nature of this pilot study and the limited data collection opportunities, the analysis focused on identifying promising trends and insights rather than establishing statistical significance. This approach is appropriate for pilot studies intended to inform future, more comprehensive research efforts (Dornyei, 2007).

FINDING AND DISCUSSION

1. Overall Transcription Success Rates

Analysis of the transcription attempts across both sessions revealed a notable improvement in students' ability to produce utterances that the STT software could accurately transcribe. In Session 1, the overall success rate was 38.4%, indicating that just over one-third of student utterances were correctly transcribed by the software. By Session 2, conducted nine weeks later, the success rate had increased to 62.1%, representing a 23.7 percentage point improvement.

This improvement was consistent across all three classes, though with some variation in the magnitude of change:

- Class 1: 36.9% (Session 1) to 58.7% (Session 2)
- Class 2: 39.2% (Session 1) to 64.8% (Session 2)
- Class 3: 39.1% (Session 1) to 62.8% (Session 2)

These improvement rates align with findings from recent controlled studies of ASR-based pronunciation training. Research by Sun (2023) with Chinese EFL learners documented similar improvement trajectories over short intervention periods, while Rogti (2025) found comparable gains among Algerian students using ASR with peer correction. The consistency of improvement across all three classes in the present study supports Liu et al.'s (2025) observation that speech recognition technology can provide reliable pronunciation feedback across various classroom implementations, even with limited technological resources.

While some of this improvement may be attributed to increased familiarity with the activity format, the consistency across all three classes suggests a genuine improvement in pronunciation intelligibility as assessed by the STT algorithm.

2. Specific Pronunciation Features

The research revealed that certain pronunciation features showed more significant improvement than others. Most notably, utterances containing /r/ and /l/ phonemes—a well-documented challenge for Japanese speakers of English (Aoyama et al., 2004)—showed substantial improvement in successful transcription, from 41.3% in Session 1 to 79.6% in Session 2.

Other features showing meaningful improvement included:

- Consonant clusters: 44.5% to 68.2% success rate
- Word-final consonants: 43.7% to 67.9% success rate
- Long/short vowel distinctions: 47.2% to 65.8% success rate

Interestingly, certain pronunciation features proved more resistant to improvement within this short intervention. For example, utterances containing "th" sounds (/θ/ and /ð/) showed relatively modest improvement (39.1% to 48.7%). This may reflect the inherent difficulty of these sounds for Japanese speakers or suggest that these features require more extensive intervention for significant improvement.

The differential improvement across various pronunciation features suggests that STT technology may be particularly effective for addressing certain phonological challenges while

being less effective for others. This finding has implications for how such technology might be strategically integrated into broader pronunciation instruction approaches.

3. Student Responses and Affective Factors

Questionnaire responses and classroom observations revealed several important affective outcomes from the STT activity. Students generally reported high levels of engagement with the activity, with 89% rating it as "fun" or "very fun" across both sessions. This positive response is particularly noteworthy given that pronunciation practice is often perceived as challenging or anxiety-provoking by EFL learners.

Several themes emerged from the qualitative data:

1) **Reduced anxiety**: Students reported feeling less anxious practicing with the technology compared to traditional pronunciation activities. One student noted, "I don't feel embarrassed when the app doesn't understand me, but I feel very nervous when people don't understand me."
2) **Concrete feedback**: The visual representation of speech as text provided clear evidence of intelligibility issues. As another student commented, "When I see the wrong words on the screen, I understand my pronunciation problem better than when my teacher just tells me I'm saying it wrong."
3) **Sense of achievement**: Successfully "beating the machine" provided immediate positive reinforcement. One student remarked, "I can see I'm getting better because the app understands me more often now."
4) **Autonomy**: Students valued the ability to make multiple attempts without time pressure. As expressed by one participant, "I can try many times and different ways to pronounce until the app understands me. In class, I only get one or two chances."
5) Recent research has increasingly recognized the importance of these affective factors in pronunciation development. Xiao (2025) demonstrated that ASR technology can significantly reduce listening and speaking anxiety while increasing flow experience among EFL learners. Similarly, Evers and Chen (2022) found that gamified ASR activities enhanced learner motivation through immediate feedback and measurable progress indicators. The positive affective outcomes observed in this pilot study corroborate these findings, suggesting that even limited ASR implementation can yield meaningful psychological benefits that support continued pronunciation practice.

These qualitative findings align with research on motivation in language learning, which emphasizes the importance of autonomy, mastery, and purpose in fostering engagement (Dörnyei, 2009). The STT activity appears to support these motivational factors by providing clear goals, immediate feedback, and a sense of progress.

4. Practical Implementation Considerations

The pilot study revealed several important considerations for implementing STT technology in EFL classrooms:

1. **Technology limitations**: Having only one device per class created a turn-taking situation that limited individual practice time. Ideally, students would have access to individual devices or work in small groups to maximize engagement.

2. **Teacher guidance**: While the software provided objective feedback on intelligibility, students often needed instructor guidance to interpret this feedback and develop effective correction strategies. The most significant improvements occurred when technology use was paired with targeted instruction on specific pronunciation features.
3. **Activity framing**: The "game" format with bonus points for successful transcription helped reduce anxiety and increase motivation. This framing was particularly important for encouraging participation from less confident students.
4. **Sentence selection**: Sentences containing specific pronunciation challenges for Japanese learners proved most effective for demonstrating improvement. The most productive practice occurred with sentences of moderate length (5-10 words) containing targeted phonological features.

These implementation considerations reflect challenges and opportunities identified in recent research on technology integration in pronunciation teaching. Inceoglu et al. (2023) emphasized the importance of instructor guidance in interpreting ASR feedback, particularly for segmental features that significantly impact intelligibility. John et al. (2025) further highlighted the value of framing ASR feedback as both corrective and confirmative, an approach that aligns with the game format employed in this study. The finding that one device per class created turn-taking limitations echoes recommendations by Liu et al. (2025) for exploring peer-collaborative approaches that maximize engagement with limited technological resources.

These findings suggest that STT technology is most effective as a complementary tool within a broader pronunciation instruction approach rather than as a standalone solution. The technology provides unique benefits in terms of objective feedback and engagement, but these benefits are maximized when integrated thoughtfully into pedagogical practice.

CONCLUSION

This pilot study demonstrates that speech-to-text technology can serve as an effective tool for pronunciation assessment and feedback in EFL contexts, even with limited implementation timeframes. The approach offers several advantages: it provides objective, immediate feedback based on intelligibility; it reduces affective barriers to pronunciation practice; and it supports learner engagement through its game-like format.

These findings contribute to a growing body of research demonstrating the effectiveness of speech-to-text technology in EFL pronunciation development. Recent systematic reviews have highlighted how ASR technology can provide objective pronunciation assessment while fostering learner autonomy (Liu et al., 2025; Inceoglu et al., 2023). The improvements observed in this pilot study, particularly for the r/l distinction, align with research by Sun (2023) showing that even short-term ASR interventions can produce measurable gains in problematic phonological features. Furthermore, the positive affective outcomes support Xiao's (2025) findings that technology-mediated pronunciation practice can reduce anxiety while maintaining engagement.

The findings suggest that STT technology is particularly effective for addressing certain segmental features, especially the /r/-/l/ distinction that is challenging for Japanese learners. The technology appears to be most beneficial when implemented with appropriate teacher guidance and integrated into a communicative language teaching approach.

Several limitations of this pilot study should be acknowledged. The limited number of sessions and the variation in student attempts make it difficult to draw definitive conclusions about

long-term effectiveness. The single-device implementation also restricted the amount of individual practice time. Additionally, as a "speaker-independent" system, the STT software may have recognized some non-standard pronunciations that human listeners would find challenging to understand, potentially overestimating intelligibility in some cases. Future research directions include exploring the efficacy of this approach with different language backgrounds, investigating the potential for more extensive implementation throughout a semester, and examining the long-term retention of pronunciation improvements gained through this method. Studies comparing single-device classroom implementation with individual practice using personal devices would also provide valuable insights for practical application.

Future implementations might explore innovative approaches recently documented in the literature. Rogti (2025) demonstrated enhanced outcomes when combining ASR with structured peer feedback, while John et al. (2025) highlighted the value of ASR for providing both corrective and confirmative feedback. Additionally, Evers and Chen (2022) showed promising results from integrating ASR within broader communicative activities that provide authentic practice contexts. These approaches could address some of the limitations identified in the present study while building on its foundational insights.

As speech recognition technology continues to improve and become more accessible, its potential as a tool for language learning expands. This research contributes to our understanding of how such technology can be effectively integrated into pedagogical practice, supporting the development of intelligible, confident English speakers in EFL contexts with practical, accessible methods that require minimal resources and preparation time.

REFERENCES

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. Journal of Phonetics, 32(2), 233-250.

Baker, A., & Burri, M. (2016). Feedback on second language pronunciation: A case study of EAP teachers' beliefs and practices. Australian Journal of Teacher Education, 41(6), 1-19.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (2001). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. Perception & Psychophysics, 61(5), 977-985.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). Teaching pronunciation: A course book and reference guide (2nd ed.). Cambridge University Press.

Dörnyei, Z. (2007). Research methods in applied linguistics. Oxford University Press.

Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei & E. Ushioda (Eds.), Motivation, language identity and the L2 self (pp. 9-42). Multilingual Matters.

Evers, K., & Chen, S. (2022). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. Computer Assisted Language Learning, 35(10), 1869-1889.

Inceoglu, S., Chen, W. H., & Lim, H. (2023). Assessment of L2 intelligibility: Comparing L1 listeners and automatic speech recognition. ReCALL, 35(1), 89-104.

John, P., Johnson, C., & Cardoso, W. (2025). Exploring automatic speech recognition for corrective and confirmative pronunciation feedback. Journal of Second Language Pronunciation.

Kartchava, E., & Ammar, A. (2014). The noticeability and effectiveness of corrective feedback in relation to target type. Language Teaching Research, 18(4), 428-452.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. Applied Linguistics, 36(3), 345-366.

Levis, J. M. (2007). Computer technology in teaching and researching pronunciation. Annual Review of Applied Linguistics, 27, 184-202.

Levis, J. M. (2018). Intelligibility, oral communication, and the teaching of pronunciation. Cambridge University Press.

Liu, Y., Rahman, F. A., & Zain, F. M. (2025). A systematic literature review of research on automatic speech recognition in EFL pronunciation. Cogent Education, 12(1).

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. Studies in Second Language Acquisition, 32(2), 265-302.

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. System, 57, 25-42.

McCrocklin, S. M. (2019). ASR-based dictation practice for second language pronunciation improvement. Journal of Second Language Pronunciation, 5(1), 98-118.

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. Foreign Language Annals, 51(3), 617-637.

Munro, M. J., & Derwing, T. M. (2015). A prospectus for pronunciation research in the 21st century: A point of view. Journal of Second Language Pronunciation, 1(1), 11-42.

Neri, A., Cucchiarini, C., & Strik, H. (2008). The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. ReCALL, 20(2), 225-243.

O'Brien, M. G., Derwing, T. M., Cucchiarini, C., Hardison, D. M., Mixdorff, H., Thomson, R. I., Strik, H., Levis, J. M., Munro, M. J., Foote, J. A., & Levis, G. M. (2018). Directions for the future of technology in pronunciation research and teaching. Journal of Second Language Pronunciation, 4(2), 182-207.

Rogti, M. (2025). Automatic pronunciation evaluation feedback and peer correction for shaping English pronunciation accuracy and interpersonal communication. Innovation in Language Learning and Teaching, 1-17.

Sardegna, V. G. (2011). Pronunciation learning strategies that improve ESL learners' linking. In J. Levis & K. LeVelle (Eds.), Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference (pp. 105-121). Iowa State University.

Sun, C. (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.

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. Applied Linguistics, 36(3), 326-344.

Ueyama, M. (2000). Prosodic transfer: An acoustic study of L2 English vs. L2 Japanese. University of California, Los Angeles.

Wallace, L. (2016). Using Google Web Speech as a springboard for identifying personal pronunciation problems. Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference, 180-186.

Xiao, Y. (2025). The impact of AI-driven speech recognition on EFL listening comprehension, flow experience, and anxiety: A randomized controlled trial. Humanities and Social Sciences Communications, 12(1), 425.