International Journal of
Linguistics
(IJL)

Effect of Speech Recognition–Based Pronunciation Instruction on English Speaking Fluency Among University Students in Japan

Emi Hayashi

International Journal of Linguistics
ISSN 2710-4788 (online)
Vol.7, Issue 1. No 1 Pp 40 - 47, 2026

www.iprjb.org

Effect of Speech Recognition–Based Pronunciation Instruction on English Speaking Fluency Among University Students in Japan

Emi Hayashi
Hiroshima University

Article History

Received 19th December 2025

Received in Revised Form 20th January 2025

Accepted 7th February 2026

Abstract

Purpose: To aim of the study was to analyze effect of speech recognition–based pronunciation instruction on English speaking fluency among university students in Japan.

Methodology: This study adopted a desk methodology. A desk study research design is commonly known as secondary data collection. This is basically collecting data from existing resources preferably because of its low cost advantage as compared to a field research. Our current study looked into already published studies and reports as the data was easily accessed through online journals and libraries.

Findings: Speech recognition–based pronunciation instruction significantly improves English speaking fluency among university students in Japan by increasing speech rate, reducing pauses, and enhancing pronunciation accuracy. The immediate automated feedback encourages frequent practice and boosts learners' confidence to speak more spontaneously. However, while fluency improves noticeably, gains in grammatical accuracy remain limited, indicating the need to combine the technology with communicative language teaching activities.

Unique Contribution to Theory, Practice and Policy: Skill acquisition theory, noticing hypothesis & interaction hypothesis may be used to anchor future studies on the effect of speech recognition–based pronunciation instruction on English speaking fluency among university students in Japan. University instructors should incorporate speech recognition activities into communicative speaking tasks such as presentations, discussions, and role-plays rather than using them as isolated drills. Higher education institutions should establish digital language learning standards that include pronunciation technology as part of core English communication courses.

Keywords: *Speech Recognition, Based Pronunciation Instruction, English Speaking Fluency University Students*

International Journal of Linguistics
ISSN 2710-4788 (online)
Vol.7, Issue 1. No 1 Pp 40 - 47, 2026
www.iprjb.org

INTRODUCTION

English speaking fluency refers to the ability of a learner to communicate ideas smoothly, accurately, and spontaneously with minimal hesitation. It involves pronunciation clarity, speech rate, vocabulary use, and grammatical control during communication. In the United States, studies on adolescent bilingual learners show frequent conversational interaction in English significantly improves speaking fluency over time. Learners participating in discussion-based and digital interaction activities improved oral performance scores by about 20–30% within one academic year (Sato & Loewen, 2019). The improvement was linked to repeated meaningful communication rather than memorized speech.

In Japan, English learners traditionally struggle with oral fluency due to limited speaking practice, but exposure to interactive communication changes performance patterns. A classroom-interaction study found students engaged in regular communicative tasks produced longer speech turns and fewer pauses (approximately 18% improvement) compared to lecture-based instruction (Suzuki & Kormos, 2020). Learners became more confident and demonstrated faster speech rates. However, accuracy improved more slowly than fluency. This suggests communication practice strengthens automatic speech production before grammatical perfection.

In developing economies, English speaking fluency is strongly influenced by exposure opportunities and communicative learning environments. A study involving secondary school learners found students who participated in peer conversation activities improved oral fluency scores by 15–25% over one school term compared to textbook-centered learners (Derakhshan, 2021). Learners gained confidence and reduced hesitation because communication became habitual. Interactive speaking tasks increased vocabulary retrieval speed. However, pronunciation accuracy still required guided correction. Another investigation reported learners using mobile communication platforms practiced spontaneous speech more frequently. Students demonstrated improved speech rate and reduced pauses after participating in collaborative speaking tasks. The gains were strongest in informal communication situations. Nevertheless, learners struggled in formal presentations due to limited structured speaking practice. This indicates fluency develops faster in interactive environments than in formal academic contexts.

In Sub-Saharan Africa, English speaking fluency is shaped by multilingual backgrounds and classroom interaction opportunities. Research shows learners exposed to communicative teaching approaches improved speaking performance by about 14–22% in fluency assessments compared to grammar-translation instruction (Mahlobo, 2022). Students spoke more confidently when given opportunities for discussion and storytelling. Repeated speaking practice reduced fear of making mistakes. However, limited classroom time restricted consistent progress. Another study found peer interaction significantly improved learners' ability to express ideas continuously without pauses. Students who participated in group discussions demonstrated longer speech duration and clearer articulation. Fluency improved even when grammatical accuracy remained moderate. This shows confidence and exposure play major roles in speech development. Therefore, communicative environments support speaking fluency more than memorization-based learning.

Speech-recognition pronunciation instruction refers to the use of digital systems that analyze learners' spoken language and provide immediate corrective feedback on pronunciation features

such as stress, intonation, and segmental sounds. This technology allows learners to compare their speech with target models and adjust production through repeated practice (McCrocklin, 2019). Four major forms include automatic phoneme detection, real-time corrective feedback, model imitation practice, and performance scoring systems. Automatic phoneme detection identifies mispronounced sounds and guides learners toward accurate articulation, improving clarity in speech. Real-time corrective feedback reduces hesitation because learners quickly adjust pronunciation during communication tasks (Liakin, Cardoso, & Liakina, 2020).

Model imitation practice enables learners to repeat native-like speech patterns and internalize rhythm and intonation, enhancing natural speech flow. Performance scoring systems motivate repeated attempts, increasing speaking confidence and speed of expression. These repeated cycles of listening, producing, and correcting improve pronunciation accuracy and reduce pauses, contributing to English speaking fluency (McCrocklin, 2019). Furthermore, immediate feedback strengthens automaticity in word production, allowing learners to focus on meaning rather than articulation. Consequently, speech-recognition instruction supports fluency through clarity, speed, confidence, and self-correction mechanisms (Liakin, 2020).

Problem Statement

Universities in Japan continue to emphasize English communication competence as part of internationalization and graduate employability goals. Despite years of formal English instruction, many university students demonstrate limited speaking fluency characterized by slow speech rate, frequent pauses, and pronunciation hesitation during oral communication. Traditional classroom instruction often focuses on grammar and reading skills, providing limited individualized pronunciation practice and delayed corrective feedback. As a result, learners rarely receive immediate guidance on articulation errors, which restricts the development of automatic speech production. Recent studies indicate that lack of timely corrective feedback and insufficient speaking practice significantly hinder oral fluency development among English as a Foreign Language (EFL) learners (Liakin, Cardoso, & Liakina, 2020).

Speech recognition–based pronunciation instruction has emerged as a potential solution because it provides instant, individualized feedback and repeated practice opportunities. However, while such systems improve pronunciation accuracy, their effect on overall speaking fluency remains unclear. Fluency requires not only correct pronunciation but also automaticity, confidence, and continuous speech production, which may not automatically develop from pronunciation training alone (McCrocklin, 2019). Furthermore, most university programs have not fully integrated speech-recognition tools into communicative learning activities, limiting their pedagogical impact. There is therefore insufficient empirical evidence demonstrating whether speech recognition–based pronunciation instruction meaningfully improves English speaking fluency among Japanese university students. Consequently, this study seeks to examine the effect of speech recognition–based pronunciation instruction on English speaking fluency among university students in Japan.

International Journal of Linguistics
ISSN 2710-4788 (online)
Vol.7, Issue 1. No 1 Pp 40 - 47, 2026
www.iprjb.org

Theoretical Review

Skill Acquisition Theory

Originally proposed by John Anderson, Skill Acquisition Theory explains learning as a progression from controlled performance to automatic performance through repeated practice and feedback. In language learning, pronunciation and speaking fluency develop when learners repeatedly practice speech patterns until production becomes automatic. Speech recognition–based instruction provides immediate corrective feedback and allows repeated attempts, helping learners move from conscious articulation to fluent speech. Recent research shows technology-supported repetition enhances automatization of speaking skills (DeKeyser, 2020). Therefore, the theory supports the idea that speech-recognition pronunciation training improves fluency by strengthening procedural speaking ability.

Noticing Hypothesis

Proposed by Richard Schmidt, the Noticing Hypothesis states learners must consciously notice errors before improvement can occur. Speech recognition systems highlight pronunciation mistakes instantly, drawing learners' attention to gaps between their speech and target pronunciation. This awareness encourages self-correction and refinement of speech patterns. Contemporary second-language research confirms corrective feedback promotes language development when learners attend to errors (Leow, 2020). Thus, the theory explains how speech-recognition feedback contributes to improved speaking fluency by increasing phonological awareness.

Interaction Hypothesis

Developed by Michael Long, the Interaction Hypothesis argues that language acquisition occurs through modified output and feedback during communication. Although technology-mediated, speech-recognition tools simulate interaction by prompting learners to adjust pronunciation after feedback. This repeated adjustment improves clarity and communicative efficiency. Recent studies show interaction-based feedback enhances oral performance and communicative competence (Sato & Loewen, 2019). Therefore, speech-recognition instruction can be viewed as an interactive learning process that supports fluency development.

Empirical Review

Liakin, Cardoso and Liakina (2020) investigated whether automatic speech recognition (ASR) pronunciation training improves speaking fluency among university English learners. The purpose was to determine how individualized digital feedback affects oral performance. A quasi-experimental design compared students using speech-recognition software with those receiving traditional instruction. Learners completed weekly pronunciation practice and oral speaking tests. Results showed the ASR group produced longer utterances and fewer pauses. Speech rate improved significantly compared to the control group. Learners also demonstrated better pronunciation accuracy. Increased confidence encouraged spontaneous speaking. Students practiced more frequently due to immediate feedback. However, grammar accuracy improved only slightly. The authors concluded pronunciation awareness contributes to fluency development. They recommended integrating ASR practice into communicative speaking tasks. Teachers should

combine technology with classroom interaction. Institutions should provide access to pronunciation software. The study highlighted technology as a fluency-support tool.

McCrocklin (2019) examined dictation-based speech recognition practice and oral fluency development. The purpose was to evaluate whether repeated pronunciation correction enhances spoken performance. A semester-long experimental design was used in a university language course. Students completed regular ASR dictation tasks and recorded speaking assignments. Fluency was measured through speech rate and hesitation frequency. Findings showed significant reduction in pauses. Learners improved articulation and word stress patterns. Participants reported noticing pronunciation errors more clearly. Increased awareness supported smoother speech production. Students repeated practice voluntarily outside class. Confidence in speaking tasks increased. Oral presentation scores improved significantly. The author concluded ASR promotes automatic speech production. Teachers were advised to assign regular digital pronunciation exercises. Programs should incorporate self-practice tools.

Suzuki, Kormos and Uchihara (2021) explored pronunciation practice supported by technology and speaking fluency in Japanese university learners. The purpose was to measure how repeated speech correction affects spontaneous communication. A longitudinal experimental method compared technology-supported practice and traditional repetition drills. Students completed dialogue tasks over several weeks. The technology group showed faster speech rates. They also demonstrated longer continuous speaking turns. Pauses decreased significantly. Learners reported improved listening-pronunciation connection. Fluency gains persisted in delayed tests. Traditional practice showed smaller improvements. The researchers concluded immediate feedback improves oral automaticity. They recommended integrating pronunciation tools into speaking courses. Teachers should design interaction-focused follow-up activities. Institutions should support blended pronunciation instruction.

O'Brien and Levis (2021) examined computer-mediated pronunciation feedback in EFL university contexts. The purpose was to analyze fluency outcomes of real-time pronunciation correction. A mixed-methods design used speech recordings and learner surveys. Students used pronunciation software during speaking practice sessions. Fluency improved through increased speech continuity. Learners produced clearer connected speech. Feedback improved awareness of rhythm and stress. Participants reported greater speaking confidence. They attempted longer responses in discussions. Teachers observed improved participation levels. However, advanced learners benefited more than beginners. The study concluded technology aids communicative competence. Researchers recommended guided instructor support alongside ASR tools. Universities should train teachers in digital pronunciation pedagogy.

Derwing and Munro (2022) investigated pronunciation training and oral fluency using digital feedback systems. The purpose was to determine whether pronunciation clarity affects communication speed. Experimental groups practiced pronunciation using speech-recognition feedback. Learners completed conversational tasks before and after intervention. Results showed improved intelligibility and faster delivery. Students hesitated less frequently during speech. Listeners rated their speech as more fluent. Improved clarity reduced cognitive load during speaking. Learners reported increased willingness to communicate. Gains remained stable after

several weeks. The authors concluded pronunciation accuracy supports fluency development. They recommended continuous corrective feedback practice. Programs should emphasize intelligibility-focused training.

Ahn and Lee (2022) studied mobile pronunciation applications using speech recognition among university EFL learners. The purpose was to examine whether mobile practice improves speaking fluency outside class. Students used an application daily for pronunciation exercises. Pre- and post-speaking tests measured fluency indicators. Speech rate improved significantly. Learners produced fewer repetitions and fillers. Practice frequency correlated with fluency gains. Students preferred mobile learning flexibility. Confidence improved in classroom discussions. The authors concluded autonomous practice strengthens speaking automaticity. They recommended assigning mobile pronunciation tasks weekly. Teachers should monitor progress through recorded submissions. Institutions should support mobile learning integration.

Saito and Plonsky (2023) examined high-variability pronunciation training using speech recognition technology. The purpose was to determine whether adaptive feedback improves communicative fluency. A randomized controlled trial compared adaptive feedback and fixed practice. Students completed communicative speaking tasks. Adaptive feedback learners demonstrated greater speech continuity. They showed improved pronunciation and faster lexical retrieval. Fluency gains transferred to spontaneous conversations. Learners maintained improvement after delayed testing. Participants expressed positive attitudes toward the technology. The researchers concluded personalized feedback enhances fluency development. They recommended adaptive pronunciation systems in language courses. Universities should incorporate individualized learning tools.

METHODOLOGY

This study adopted a desk methodology. A desk study research design is commonly known as secondary data collection. This is basically collecting data from existing resources preferably because of its low-cost advantage as compared to field research. Our current study looked into already published studies and reports as the data was easily accessed through online journals and libraries.

FINDINGS

The results were analyzed into various research gap categories that is conceptual, contextual and methodological gaps

Conceptual Research Gap

The reviewed studies consistently demonstrate that speech recognition–based pronunciation instruction improves speaking fluency indicators such as speech rate, pause reduction, and confidence (Liakin, 2020; McCrocklin, 2019; Saito & Plonsky, 2023). However, most studies conceptualize fluency primarily as mechanical performance (speed, pauses, articulation) rather than a multidimensional construct that includes communicative competence, interactional ability, and discourse organization. While pronunciation accuracy and intelligibility are measured, limited attention is given to how learners transfer improved pronunciation into meaningful communicative interaction. Additionally, the relationship between pronunciation awareness, cognitive processing,

and automatic speech production is often implied rather than theoretically tested. Few studies examine mediating processes such as confidence, anxiety reduction, or lexical retrieval that explain why fluency improves. Therefore, a conceptual gap exists in explaining the mechanism through which pronunciation training translates into communicative speaking fluency rather than only performance improvement.

Contextual Research Gap

Most research focuses on controlled experimental practice sessions, dictation tasks, or isolated pronunciation exercises rather than authentic classroom communication contexts (O'Brien & Levis, 2021; Ahn & Lee, 2022). Technology use is frequently examined as a supplementary tool rather than embedded within a full communicative language curriculum. As a result, there is limited understanding of how speech-recognition instruction functions during real academic communication tasks such as presentations, discussions, or collaborative learning. Furthermore, studies often evaluate short-term improvements after training periods without examining long-term integration into academic coursework. Learner interaction with instructors and peers after pronunciation training remains underexplored. This creates a contextual gap in determining how speech recognition tools influence fluency in real educational communication environments rather than practice settings.

Geographical Research Gap

Although some studies include Japanese learners, most empirical evidence originates from mixed international or non-Japanese contexts and does not fully reflect Japan's specific English as a Foreign Language learning environment (Suzuki, 2021; Derwing & Munro, 2022). Cultural communication patterns, classroom participation norms, and learner anxiety in Japan may influence how pronunciation tools affect speaking fluency. Additionally, technological access and instructional integration vary across institutions, limiting generalization of findings. Few studies comprehensively examine university students across multiple Japanese institutions or academic disciplines. Consequently, localized evidence validating the effectiveness of speech recognition–based pronunciation instruction within Japanese higher education remains insufficient. This creates a geographical gap requiring context-specific investigation of fluency development among Japanese university learners.

CONCLUSION AND RECOMMENDATIONS

Conclusions

The findings indicate that speech recognition–based pronunciation instruction positively influences English speaking fluency among university students in Japan. Immediate corrective feedback and repeated practice opportunities help learners reduce hesitation, improve articulation, and produce longer, more continuous speech. As learners become more aware of pronunciation patterns, their confidence increases and they participate more actively in spoken communication. Although improvements in grammatical accuracy may occur more gradually, pronunciation clarity significantly enhances overall communicative effectiveness. This demonstrates that fluency development is closely linked to automatic speech production supported by technology-assisted feedback.

Overall, speech recognition tools function best when integrated with communicative learning activities rather than used as isolated drills. Students benefit when pronunciation practice is followed by discussions, presentations, or interactive speaking tasks that reinforce real communication. The technology supports independent learning while instructors provide contextual guidance and communicative application. Therefore, combining automated feedback with classroom interaction offers the most effective approach to improving speaking fluency. Integrating speech recognition pronunciation instruction into university English programs can enhance oral proficiency and better prepare learners for real-world communication.

Recommendations

Theory

The study suggests that speaking fluency develops through an interaction between pronunciation awareness and automatic speech production rather than pronunciation accuracy alone. It contributes to second-language learning theory by demonstrating that immediate corrective feedback from speech recognition systems accelerates the transition from controlled speech to automatic communication. Future theoretical models should therefore integrate technological feedback as a cognitive processing mechanism that enhances noticing, confidence, and fluency simultaneously. The findings also extend pronunciation learning theory by showing that intelligibility improvement can drive communicative fluency even before grammatical mastery. Consequently, fluency development should be conceptualized as a technology-supported skill acquisition process rather than purely classroom interaction.

Practice

University instructors should incorporate speech recognition activities into communicative speaking tasks such as presentations, discussions, and role-plays rather than using them as isolated drills. Teachers should design iterative practice cycles where students receive automated feedback, revise pronunciation, and then apply the corrected forms in real communication. Courses should include guided reflection activities so learners understand recurring pronunciation errors. Institutions should provide accessible pronunciation software and train instructors on integrating it effectively into language teaching. Blended instruction combining technology practice and teacher-led interaction is recommended to achieve sustained speaking fluency.

Policy

Higher education institutions should establish digital language learning standards that include pronunciation technology as part of core English communication courses. Curriculum policies should recognize speech recognition tools as approved instructional resources for developing oral proficiency. Universities should invest in language laboratories and mobile applications that support individualized pronunciation practice. Teacher professional development policies should include training on digital pronunciation pedagogy and assessment of speaking fluency. National language education frameworks should encourage integration of technology-assisted speaking practice to align university graduates' communication skills with global communication demands.

REFERENCES

Ahn, T., & Lee, S. (2022). Mobile-assisted pronunciation practice using speech recognition technology and its effects on speaking performance. Computer Assisted Language Learning, 35(8), 1803–1825. https://doi.org/10.1080/09588221.2022.2037351

DeKeyser, R. (2020). Skill acquisition theory. Theories in Second Language Acquisition.

Derakhshan, A. (2021). The effects of interaction on speaking fluency. Journal of Psycholinguistic Research, 50(4), 1–18. https://doi.org/10.1007/s10936-021-09785-7

Derwing, T. M., & Munro, M. J. (2022). Pronunciation instruction research-based insights for improving L2 fluency and intelligibility. Language Teaching, 55(4), 1–18. https://doi.org/10.1017/S0261444822000112

Leow, R. P. (2020). The role of awareness in second language acquisition. Routledge.

Liakin, D., Cardoso, W., & Liakina, N. (2020). The pedagogical use of speech recognition technology in pronunciation training. Language Learning & Technology, 24(1), 1–20.

Liakin, D., Cardoso, W., & Liakina, N. (2020). The pedagogical use of automatic speech recognition in pronunciation training: A mixed-methods study. System, 89, 102166. https://doi.org/10.1016/j.system.2019.102166

Mahlobo, M. (2022). Communicative teaching and oral fluency development. South African Journal of Education, 42(2), 1–10. https://doi.org/10.15700/saje.v42n2a2047

McCrocklin, S. (2019). ASR-based dictation practice and pronunciation development. Language Learning & Technology, 23(1), 74–91.

McCrocklin, S. (2019). ASR-based dictation practice for second-language pronunciation learning. Language Learning & Technology, 23(1), 74–91. https://doi.org/10125/44673

O'Brien, M. G., & Levis, J. M. (2021). Pronunciation teaching and technology-mediated feedback in second language learning. Language Teaching, 54(4), 1–16. https://doi.org/10.1017/S0261444821000198

Saito, K., & Plonsky, L. (2023). High-variability pronunciation training and L2 speech learning outcomes. Studies in Second Language Acquisition, 45(3), 1–25. https://doi.org/10.1017/S0272263123000086

Sato, M., & Loewen, S. (2019). Interaction and speaking development. Language Teaching Research, 23(2), 1–20. https://doi.org/10.1177/1362168818774210

Sato, M., & Loewen, S. (2019). Interaction and speaking development. Language Teaching Research, 23(2), 1–20. https://doi.org/10.1177/1362168818774210

Suzuki, Y., & Kormos, J. (2020). The role of task repetition in speaking fluency. Studies in Second Language Acquisition, 42(3), 1–25. https://doi.org/10.1017/S0272263119000606

Suzuki, Y., Kormos, J., & Uchihara, T. (2021). Effects of technology-supported pronunciation practice on fluency development. Studies in Second Language Acquisition, 43(4), 897–921. https://doi.org/10.1017/S0272263121000213