The Implementation of Automated Speech Recognition (ASR) in ELT Classroom: A Systematic Literature Review from 2012-2023

*1Ngadiso, 1Alan Surya Wijaya, 1Alizta Marduama, 1Almas Khoirun Nisa, 1Aurismadiva Nur Fahmi
1Universitas Sebelas Maret, Indonesia

*Correspondence:
ngadisodok@staff.uns.ac.id

Abstract

This study systematically reviews the literature on the use of Automated Speech Recognition (ASR) in English Language Teaching (ELT) classrooms, focusing on its application in speaking assessments. ASR technology, which converts speech audio into text, offers significant advantages in teaching pronunciation and evaluating learners' spoken language. It opens avenues for interactive aural engagements between students and computers. This research aims to scrutinize the effectiveness of ASR in speaking assessment, emphasizing the validity of automated scores and addressing potential technological inaccuracies or limitations. Despite existing research on technology's role in speaking evaluation, there remains a paucity of empirical studies that systematically review ASR's educational applications. This gap underscores the need for an up-to-date, comprehensive exploration of ASR's implementation in education. For this review, ten studies from 2012 to 2023 were selected from renowned academic databases, including Taylor & Francis, Wiley, and Springer. The findings highlight ASR's educational benefits, such as enhancing student progress, fostering interaction, and offering significant pedagogical contributions. Notably, the integration of ASR in speaking assessments facilitates a synergistic relationship between human and automated scoring, enriching the assessment process. This study serves as a valuable resource for educators seeking to optimize speaking assessments in ELT classrooms through ASR technology. It sheds light on the methodological advancements and pedagogical implications of integrating ASR, providing a foundation for future research and practical applications in language education.

Keywords: Speaking assessment, Automated Speech Recognition (ASR), speech technology, speaking proficiency, oral communication

INTRODUCTION

The exploration of Automated Speech Recognition (ASR) technology has formed a substantial part of recent academic inquiry. ASR, a technology that interprets spoken

language into text, has seen remarkable advancements in recent years (Ivanov et al. 2015). This progression has led to its extensive application in various commercial systems, including travel reservations, financial services, and weather forecasting, as highlighted by Godwin (2009). Despite its technological advancements, ASR's integration into English Language Teaching (ELT) has been met with mixed reactions. The field has witnessed the emergence of numerous commercial English language learning software products, often falling short of their ambitious promises. This inconsistency in efficacy has led to a turbulent reputation for ASR within the ELT community. Historically, ASR has been a component of computer-assisted language learning (CALL) systems, also referred to as "voice-interactive CALL" (Carrier, 2017). These systems primarily utilized ASR for pronunciation training and initiating dialogic interactions, underscoring its potential in language education. Beyond pronunciation aid, ASR offers broader applications in analyzing learners' speech, fostering opportunities for interactive auditory exchanges between students and computer systems.

The transformative aspect of ASR lies in its ability to convert speech audio into written text, thus serving as a critical tool for teaching pronunciation (Inceoglu, 2020). Its use extends to evaluating learners' speech on a wider scale and establishing foundations for auditory interactions. ASR facilitates language learning by enabling programs to "listen" to learners' pronunciation, offering formative assessment and feedback on phonological accuracy. A key aspect of ASR's role in ELT is its contribution to pedagogical practices. It enables instructors to effectively assess spoken English in classrooms, bridging the gap between technology and teaching. The technology's impact on pedagogy extends to how educators perceive its benefits and limitations in second language (L2) instruction. ASR's capability to offer direct feedback and pronunciation assessment simplifies the teaching process, making it an invaluable tool in modern language education.

The assessment of English-speaking skills through Automated Speech Recognition (ASR) and Spoken Dialogue Systems (SDS) has been a focus of numerous studies in recent years, as evidenced by works such as Laughlin et al. (2020), Bashori et al. (2022), Gu et al. (2020), and Koizumi (2020). ASR, forming the basis of interactive SDS, incorporates both speech and natural language processing technologies to enable comprehensive human-machine interactions. This integration has proven instrumental in providing students with targeted and meaningful speaking practice. A notable trend over the past decade is the accelerating advancement in the use of ASR and SDS for assessing English speaking skills. Earlier, the application of these technologies in language assessment was limited, with ASR being more oriented towards computer-assisted language learning (CALL) rather than evaluative grading, and SDS focused on development in specific application domains (Litman et al., 2018). However, their role in speaking assessments has significantly expanded recently.

Several studies, such as those by Laughlin et al. (2020), have highlighted the effectiveness of using automated voice recognition and natural language processing in engaging students in productive speaking practice, particularly in flipped classroom models. Research by Bashori et al. (2022) demonstrated that ASR-based websites substantially aid students in enhancing their receptive vocabulary, showcasing the technology's capability to improve both vocabulary and pronunciation. Despite these advancements, ASR's journey in English Language Teaching (ELT) has been tumultuous.

Godwin (2009) notes its widespread application in various commercial systems, yet its implementation in ELT has often been plagued by overpromising and underdelivering commercial language learning software. Earlier applications of ASR in CALL, as described by Carrier (2017), were primarily focused on pronunciation training and initiating dialogues rather than comprehensive language assessment. Moreover, the reliability of ASR in ELT suffered due to high rates of false positives and negatives, leading to frustrating learning experiences. However, recent technological advancements have significantly improved recognition quality and accuracy, renewing interest and expanding possibilities in language learning and assessment.

Despite extensive research on Automated Speech Recognition (ASR) and Spoken Dialogue Systems (SDS) in language education, a significant gap persists. A key issue is the frequent development of technology-based speech recognition tools without adequate teacher involvement. The complexity of ASR is amplified by factors such as background noise, low signal-to-noise ratio, pronunciation variations, and the continuous nature of speech, which are particularly challenging when processing non-native speech (Litman et al., 2018; Cucchiarini et al., 2010). Moreover, studies have primarily focused on constrained speech tasks, such as reading aloud or elicited imitation, which do not fully capture the nuances of spontaneous speech. Attempts to assess more natural, less controlled speech have been made (e.g., Cuccharini et al., 2002; Zechner et al., 2009), but these present greater challenges for speech technology, resulting in increased errors. The calibration of task difficulty also remains a concern, as overly simple tasks may not generate sufficient errors for meaningful analysis. In addition, the deficiency in the existing literature is the limited integration of systematic research in evaluating English as a Foreign Language (EFL) speaking skills using advanced speech technologies like ASR and SDS, especially in diverse educational settings and among students with varying abilities, including those with disabilities. This gap highlights the need for an in-depth examination of ASR's role in education through systematic literature reviews.

In addressing the identified gaps, this study embarks on a comprehensive and updated systematic literature review of Automated Speech Recognition (ASR) implementation in educational contexts. Central to this investigation are two critical research objectives. Firstly, the study seeks to unravel the current research status of ASR in education. This involves a thorough exploration of existing literature to gauge the depth and breadth of ASR's application in various educational settings. Secondly, the study aims to uncover the benefits and practical implications of applying ASR technology in the realm of language assessment. Special emphasis is placed on its impact on enhancing students' speaking skills. By achieving these objectives, the study intends to offer valuable insights for educators and students considering the use of ASR in speaking assessments. It is particularly focused on evaluating the balance between human and automated scoring, assessing their validity, and identifying potential challenges or errors associated with the use of ASR technology in language assessment.

METHOD

To conduct a rigorous systematic review on the topic of Automated Speech Recognition (ASR) in education, we meticulously followed the three-step framework proposed by Mcquade (2021) and Tranfield et al. (2003). This comprehensive approach

consists of planning the review, conducting the review, and reporting and disseminating the findings. In the initial stage of our research, we established the need and objectives for this systematic literature review. This was accomplished by conducting a scoping study in the relevant field of speech technology in education, which helped us in identifying the key areas that required in-depth exploration. This stage was critical in laying the groundwork and defining the scope of our review.

Furthermore, our focus shifted towards the meticulous selection of high-quality research papers. We conducted an extensive keyword search across several acclaimed academic databases, including Taylor & Francis, Wiley Library, and Springer. The selection of these databases was strategic; our aim was to source authoritative research articles that were suitable for in-depth analysis and to ensure that our review would yield representative and reliable conclusions. To guide our selection, we employed specific inclusion criteria, focusing on papers that mentioned key terms such as "ASR", "speech technology," "technology in speaking assessment," "pronunciation," "speaking proficiency," "oral communication," and "automated speech recognition." Moreover, we imposed temporal and linguistic boundaries on our search, limiting it to papers published within the last decade (2012–2022) and ensuring that all selected articles were written in English.

To ensure a comprehensive capture of articles relevant to our study, we utilized a set of representative search terms that accurately reflected the essence of our research. These keywords were organized into two main clusters: "Automated Speech Recognition" and "Secondary School." The inclusion of related terms such as "middle school", "high school," and "children" was crucial for gathering information pertinent to our target age groups. Our initial search, using the string "Automated Speech Recognition" AND (middle school OR high school OR secondary school OR primary school OR elementary school OR kindergarten OR preschool OR children OR child), was conducted within the Social Sciences Citation Index (SSCI) of the chosen databases. This search yielded a total of 20 papers. We then meticulously examined the titles, abstracts, and methods sections of these papers against our predefined inclusion criteria. Through this careful screening, we narrowed down our selection to 10 papers that met our rigorous standards for further in-depth analysis.

FINDING AND DISCUSSION
RQ1: What is the current research status of ASR in the ELT Classroom?

We discovered information regarding the current research on Automated Speech Recognition (ASR) in ELT classrooms or speaking classrooms.
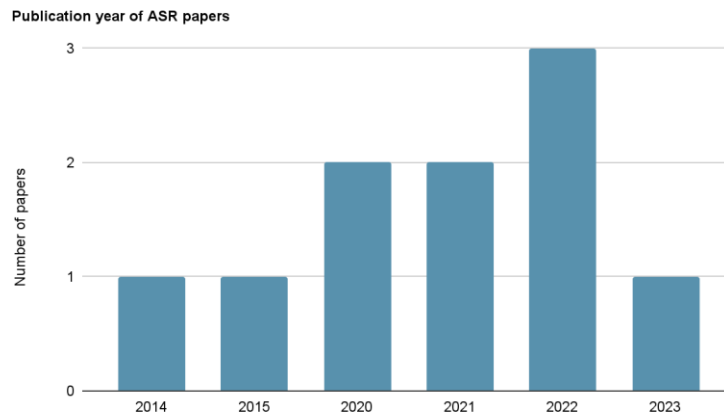
Publication year of ASR papers

Figure 1. Publication year of ASR papers

The distribution of the selected empirical studies, as illustrated in Figure 1, offers a revealing perspective on the recent trends in research related to ASR in the educational sector. Among the 10 chosen studies, there is a noticeable concentration of publications post-2014, underscoring a growing academic interest in this field. This trend is particularly evident with the emergence of three articles in 2022 alone, signaling a significant surge in well-founded empirical research in the last five years. This increase can be attributed to advancements in technology, particularly in areas that require computers to interpret, understand, and respond to user voice commands. These developments have spurred a heightened focus on enhancing human-machine interaction. A critical aspect of this burgeoning research area, as noted by Kanabur et al. (2019), is the understanding of human speech production and the development of accurate speech recognition models. This focus is not only fundamental to the technological advancement of ASR systems but also pivotal to their effective application in educational settings. Given the rapid progression in empirical research and the expanding potential for ASR's applicability in education, it is anticipated that the coming years will witness a continued upsurge in evidence-based research in this domain. Such advancements are poised to further transform the landscape of educational technology, particularly in the realm of language learning and assessment.
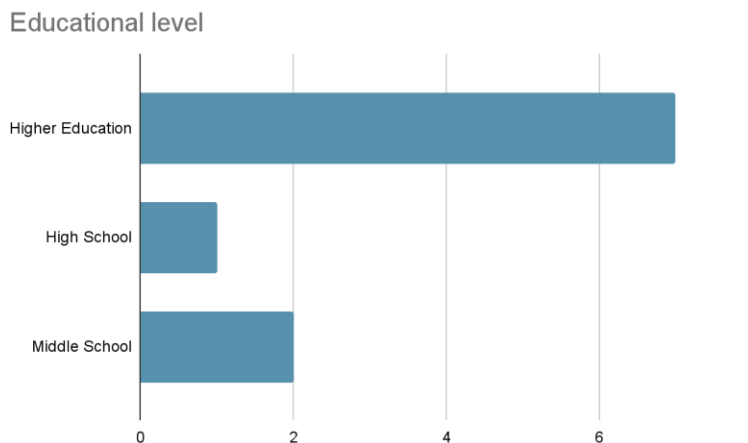
Educational level

Figure 2. The educational level of ASR papers

A significant number of studies within the realm of Automated Speech Recognition (ASR) in English Language Teaching (ELT) classrooms have predominantly centered on secondary and higher education students. This trend points to an imbalanced and often overlooked distribution across different age cohorts in ASR research. The emphasis on integrating ASR technologies in advanced educational settings, such as universities or sophisticated language courses, has highlighted the need for a deeper comprehension of these technologies and their practical applications in language learning environments. Students in higher education typically have more advanced expectations regarding the refinement and effectiveness of ASR tools.

This demand necessitates the development of more intricate features and functionalities within ASR systems to adequately meet their language learning needs (Knill et al. 2018). Furthermore, the use of ASR in these settings often requires educators to possess a specialized understanding of how to effectively incorporate this technology into their teaching methods. This situation underscores the importance of tailored professional development programs designed to equip educators with the necessary skills and knowledge for successfully integrating ASR technology in higher education environments. Understanding the unique requirements and challenges inherent at this educational level is crucial. It provides critical insights into how ASR technology can be optimized for use in higher education contexts, ensuring that it meets the specific needs of both learners and educators. Recognizing and addressing these aspects is essential for the effective assimilation and utilization of ASR in advanced educational settings, ultimately contributing to more effective and engaging language learning experiences.

RQ2: What are the benefits of applying ASR in education?

The application of Automated Speech Recognition (ASR) in English Language Teaching (ELT) classrooms offers a range of significant benefits.

Author(s)	Categories	Sub-Categories
Ahn, T.Y. and Lee, S. (2015)	Students' progress	Provide positive attitude Improve speaking skills Increases confidence Improves motivation
Bashori, M., Van, H. R., Strik, H., & Cucchiarini, C. (2022)	Students' progress	Improve vocabulary skill Improve pronunciation skill
Gu, L., Davis, L., Tao, J., & Zechner, K. (2020)	Pedagogical contributions	Provide immediate feedback Provide self-paced learning Create a flexible learning environment
Koizumi, R. (2022)	Interaction	Provide interaction opportunities (student-teacher) Provide interaction opportunities (student-student) Provide interaction opportunities (student-material)
Laughlin, V.T., Sydorenko, T., & Phoebe, D. (2020)	Pedagogical contributions	Enables meaningful learning Enables the teacher to be a facilitator Provides collaboration opportunities

Kholis, A. (2021)	Improved Pronunciation Skills	Enhance accuracy in pronunciation Increase ability to identify and correct pronunciation errors Improve fluency and naturalness in spoken language
	Motivation and Engagement	Increase student motivation to practice pronunciation Enhance engagement in pronunciation learning activities Gamify features and interactive nature of ASR tools
	Individualized Learning	Personalized feedback and tailored instruction Self-paced learning opportunities Adaptation to individual pronunciation needs
	Error Identification and Correction	Accurate identification of pronunciation errors Immediate feedback on pronunciation mistakes Targeted guidance for error correction
	Supplementing Language Teaching	Additional resources for pronunciation instruction Support for teachers in providing pronunciation feedback Integration of technology in language teaching methodologies
Chan, D. M. & Ghosh, S. (2022)	Performance Improvement	Accurate transcription Enhance Language understanding
	Robustness and Adaptability	Noise Handling Speaker Variation Adaptation Environment Adaptation
	Disentangled Representations	Content-encoding Focus Context-encoding Separation
	Generalization Enhancement	Reduce Spurious Correlations Improve Adaptability to Speaking Styles and Accents
	Semi-Supervised Learning	Learning from Limited-Labeled Data Reduce Dependency on Extensive Datasets Better Performance in Noisy Environments
Papi, S., Gretter, R., Matassoni, M., & Falavigna, D. (2021)	Improved Noisy Situation Handling	Reduction in Substitutions and Deletions Real-time feedback on pronunciation and grammar aids language acquisition
	Enhanced Language Learning Process	Personalized learning caters to individual student needs. Automated grading reduces instructor workload.

		Timely feedback helps students track progress.
	Efficient Evaluation and Feedback	Diverse neural network analysis enhances understanding of proficiency levels. Enable customized teaching approaches.
	Tailored Learning Strategies	Identify specific learning challenges and strengths. Support targeted interventions for each student.
	Personalized Learning Paths	Integrates with language learning software for interactive exercises. Promote active student engagement.
	Enhanced Computer-Aided Learning Real-Time Progress Monitoring	Continuous assessment of spoken proficiency. Facilitate timely intervention for learning gaps.
Liu, Y., Li, Y., Deng, G., Juefei-Xu, F., Du, Y., Zhang, C., Liu, C., Li, Y., Ma, L., & Liu, Y. (2023)	Valid Test Cases	Simulate real-life stuttering speech. More accurate evaluation of ASR systems
	Exposing Errors	Highlight the limitations and weaknesses of ASR systems in recognizing and transcribing stuttering speech patterns.
	Similar Stuttering Audio	The evaluation of ASR systems using ASTER is accurate and reliable.
	Bug Identification	Categorize recognition errors in ASR systems into five bug types. Identify specific areas of improvement for ASR systems.
Tao, J., Evanini, K., & Wang, X. (2014)	Enhanced automated scoring	Calculate more precise scores based on the extracted features.
	Efficient assessment process	More reliable and consistent assessment of students' spoken language proficiency. Reduce the need for manual transcription and scoring. Automatically transcribe and score spoken responses.
	Improved accuracy	Accurately transcribe speech from non-native speakers.

The integration of Automatic Speech Recognition (ASR) technology into language learning has brought about a host of significant benefits, transforming traditional language instruction into a more interactive, self-paced, and personalized experience. ASR tools have significantly enhanced students' progress by offering immediate feedback and personalized learning paths. This technology facilitates interactive opportunities between students, teachers, and learning materials, contributing to improved pronunciation skills and vocabulary acquisition. Studies like those by Ahn & Lee (2015) and Bashori et al. (2022) have shown how ASR-based language learning systems, such as I Love Indonesia (ILI) and NovoLearning (NOVO), significantly improve vocabulary and pronunciation skills among

Indonesian secondary school students. These advancements have led to increased student engagement and motivation, making language learning more appealing and effective.

The adaptability and robustness of ASR technology, including its ability to handle noise and separate context in encoding, has improved the accuracy of pronunciation assessment and feedback. This precision allows for more efficient evaluation and tailored interventions. ASR's role in enhancing generalization through diverse neural network analyses has also enabled a deeper understanding of students' proficiency levels, leading to customized teaching approaches (Chen, 2022). Scenario-based tasks have further encouraged student engagement through contextualized language learning, providing significant opportunities for authentic speaking practice.

Furthermore, Research by Gu et al. (2020) demonstrates the value of computer-based speaking tests in providing diagnostic feedback for test preparation, such as for the TOEFL iBT examination. Additionally, studies by Koizumi (2022) and Timpe-Laughlin et al. (2020) have explored the pedagogical perspective, investigating teachers' perceptions of the benefits and limitations of ASR technology in L2 instruction and its implementation in specific teaching contexts. However, despite its potential, the implementation of ASR systems has encountered challenges, particularly in recognizing and transcribing stuttering speech patterns. Initiatives like the Assessment of Stuttering Speech with Error Recognition (ASTER) aim to provide accurate evaluations and insights into the recognition of errors, leading to targeted improvements.

In addition, the early attempts to apply Automatic Speech Recognition (ASR) to language acquisition encountered mixed results, primarily due to the use of systems designed for native speakers and other purposes, such as dictation. Early studies by Coniam (1999) and Derwing et al. (2000) highlighted these challenges. The difficulty in ASR for native speech arises from factors like background noise, low signal-to-noise ratio (SNR), disfluencies, pronunciation variation, and the blending of terms in speech. However, ASR for learner's speech, particularly in second language (L2) learning, poses even greater challenges. The grammar, vocabulary, and pronunciation in L2 speech often deviate significantly from native speech norms, impacting all three "knowledge sources" of the ASR system. L2 speech also contains more disfluencies and hesitation phenomena, which vary with the learner's proficiency level (Cucchiarini et al., 2010).

These disparities between first language (L1) and L2 speech significantly impair ASR performance. As noted by Van (2001) and Van et al. (2010), the performance issues in early ASR applications were more due to the inappropriate use of systems rather than inherent flaws in ASR technology itself. Recognizing this, professionals in the field have emphasized the need to tailor ASR systems to the specific needs of language learners and to integrate student-specific materials in the learning process. The core of ASR systems lies in training language and acoustic models. These systems utilize a lexicon and a vast corpus of audio files as input. The language model calculates probabilities for individual words and their sequences, while acoustic models represent the sounds of a language. The lexicon bridges the acoustic and language models, providing both phonological and orthographic transcriptions for each entry. Lexicons often include multiple entries for words to account for different pronunciation variants.

Once trained, ASR systems can process new speech samples, employing various knowledge sources to recognize spoken words accurately (Malik et al. 2021). This process

underlines the importance of customizing ASR systems to the specific context of L2 learning, taking into account the unique linguistic characteristics and challenges faced by language learners. As the technology and understanding of ASR's applications in language learning have evolved, there has been a shift towards developing more learner-centric ASR systems that are better suited to the needs and nuances of L2 speech. The efficiency of ASR technology in assessment processes has enabled more precise scoring and reduced the need for manual transcription, saving time and resources for both teachers and students. In conclusion, ASR technology has demonstrated its potential in fostering interactive and personalized learning experiences. It has transformed traditional language instruction into a more self-directed and guided learning process. While addressing existing challenges remains crucial, the overall benefits of ASR in educational settings support the premise that this technology can significantly aid students' progress, make pedagogical contributions, and facilitate enhanced student-teacher interactions.

CONCLUSION

This systematic literature review was conducted with the aim of exploring the gap in the application of Automatic Speech Recognition (ASR) technology in language assessment, particularly focusing on the balance between human and automated scoring based on their validity and the potential challenges or errors that may arise when using ASR as a tool for speaking assessment. The review revealed several key factors that highlight the benefits of using ASR in assessing English speaking skills. These include the enhancement of students' progress, increased interaction, significant pedagogical contributions, and improved performance accuracy. One of the critical findings is the role of pedagogical contributions in validating the scores obtained from ASR. These contributions are seen in the effective collaboration between human and automated scoring, ensuring a more comprehensive and accurate assessment of students' speaking abilities.

An important observation from the review is that significant problems did not emerge during the application of ASR in classroom settings. This indicates that, when implemented effectively, ASR technology can be a valuable tool in assessing speaking abilities without major disruptions or issues. This aspect is particularly beneficial for teachers and students in secondary school English as a Foreign Language (EFL) classes, where the need for accurate and efficient assessment of speaking skills is paramount. Overall, the findings from this systematic literature review provide insightful implications for the use of ASR technology in language learning and assessment. They suggest that ASR, when used thoughtfully and in conjunction with traditional pedagogical methods, can significantly enhance the language learning process, particularly in speaking assessments. This integration of ASR in EFL classrooms can offer teachers and students a more dynamic and effective approach to language assessment, contributing to improved learning outcomes and language proficiency.

REFERENCES

Ahn, T. Y., & Lee, S. M. (2015). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology, 47*(4), 778–786. https://doi.org/10.1111/bjet.12354

Bashori, M., Van, H. R., Strik, H., & Cucchiarini, C. (2022). "Look, I can speak correctly": learning vocabulary and pronunciation through websites equipped with automatic speech recognition technology. *Computer Assisted Language Learning,* 1–29. https://doi.org/10.1080/09588221.2022.2080230

Carrier, M. (2017). Automated Speech Recognition in language learning: potential models, benefits and impact. *Training, Language and Culture, 1*(1), 46-61. http://doi.org/10.29366/2017tlc.1.1.3

Chan, D. M. & Ghosh, S. (2022). Content-context factorized representations for automated speech recognition. *Interspeech 2022*. Doi: https://doi.org/10.48550/arXiv.2205.09872

Chen, Y. C. (2022). Effects of technology-enhanced language learning on reducing EFL learners' public speaking anxiety. Computer Assisted Language Learning, 1-25. https://doi.org/10.1080/09588221.2022.2055083

Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. *System, 27*(1), 49–64. https://doi.org/10.1016/S0346-251X(98)00049-9

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech*. Journal of the Acoustical Society of America, 111*(6), 2862–2873. https://doi.org/10.1121/1.1471894

Cucchiarini, C., Van, D. J., & Strik, H. (2010). Fluency in non-native read and spontaneous speech. *Proceedings of the DiSS-LPSS Joint Workshop 2010*, 15–18. Tokyo, Japan: University of Tokyo.

Derwing, T. M., Munro, M. J. & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly, 34*(3), 592–603. https://doi.org/10.2307/3587748

Gu, L., Davis, L., Tao, J., & Zechner, K. (2020). Using spoken language technology for generating feedback to prepare for the Toefl ibt® test: A user perception study. *Assessment in Education: Principles, Policy, and Practice, 28*(1), 58–76 https://doi.org/10.1080/0969594x.2020.1735995

Inceoglu, S., Lim, H., & Chen, W. H. (2020). ASR for EFL Pronunciation Practice: Segmental Development and Learners' Beliefs. Journal of Asia TEFL, 17(3), 824. http://dx.doi.org/10.18823/asiatefl.2020.17.3.5.824

Ivanov, A. V., Ramanarayanan, V., Suendermann-Oeft, D., Lopez, M., Evanini, K., & Tao, J. (2015). Automated Speech Recognition Technology for dialogue interaction with Non-Native interlocutors. *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. https://doi.org/10.18653/v1/w15-4617

Godwin-Jones, R. (2009). Speech tools and technologies.

Kanabur, V. & Harakannanavar, S.S. (2019). An extensive review of feature extraction techniques, challenges and trends in automatic speech recognition. *Image, Graphics and Signal Processing, 5*, 1-12. Doi: https://doi.org/10.5815/ijigsp.2019.05.01

Kholis, A. (2021). Elsa speak app: Automatic speech recognition (ASR) for supplementing english pronunciation skills. *Pedagogy: Journal of English Language Teaching, 9*(1). Doi: https://doi.org/10.32332/pedagogy.v8i1

Koizumi, R. (2022). L2 speaking assessment in secondary school classrooms in Japan. Language *Assessment Quarterly, 19*(2), 142–161. https://doi.org/10.1080/15434303.2021.2023542

Knill, K., Gales, M., Kyriakopoulos, K., Malinin, A., Ragni, A., Wang, Y., & Caines, A. (2018, September). Impact of ASR performance on free speaking language assessment. In Interspeech 2018 (pp. 1641-1645). International Speech Communication Association (ISCA). https://doi.org/10.21437/interspeech.2018-1312

Laughlin, V. T., Sydorenko, T., & Daurio, P. (2020). Using spoken dialogue technology for L2 speaking practice: what do teachers think*? Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2020.1774904

Litman, D., Strik, H., & Lim, G. S. (2018). Speech technologies and the assessment of second language speaking: approaches, challenges, and opportunities. *Language Assessment Quarterly, 15*(3), 294–309. https://doi.org/10.1080/15434303.2018.1472265

Liu, Y., Li, Y., Deng, G., Juefei-Xu, F., Du, Y., Zhang, C., Liu, C., Li, Y., Ma, L., & Liu, Y. (2023). ASTER: Automatic speech recognition system accessibility testing for stutterers. Doi: https://doi.org/10.48550/arXiv.2308.15742

Malik, M., Malik, M. K., Mehmood, K., & Makhdoom, I. (2021). Automatic speech recognition: a survey. Multimedia Tools and Applications, 80, 9411-9457. https://link.springer.com/article/10.1007/s11042-020-10073-7

Mcquade, K. E., Harrison, C., & Tarbert, H. (2021). Systematically reviewing servant leadership. European Business Review, 33(3), 465-490. https://doi.org/10.1108/EBR-08-2019-0162

Papi, S., Gretter, R., Matassoni, M., & Falavigna, D. (2021). Mixtures of deep neural experts for automated speech scoring. *Proceedings of Interspeech 2020*. Doi: https://doi.org/10.21437/Interspeech.2020-1055

Paul, D., & Parekh, R. (2011). Automated speech recognition of isolated words using neural networks. *International Journal of Engineering Science and Technology (IJEST)*, 3(6), 4993-5000.

Strik, H., Neri, A., & Cucchiarini, C. (2008). Speech technology for language tutoring. *Proceedings of LangTech 2008*, 73–76.

Tao, J., Evanini, K., & Wang, X. (2014). The influence of automatic speech recognition accuracy on the performance of an automated speech assessment *system. EEE Spoken Language Technology Workshop (SLT)*, pp. 294-299. Doi: https://doi.org/10.1109/SLT.2014.7078590

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management, 14*(3), 207–222. Doi: http://doi.org/10.1111/1467-8551.00375

Van, C. D. (2001). Recognizing speech of goats, wolves, sheep, and non-natives. *Speech Communication, 35*(1-2), 71–79. https://doi.org/10.1016/S0167-6393(00)00096-0

Van, D. J., Cucchiarini, C., & Strik, H. (2010). Optimizing automatic speech recognition for low-proficient non-native speakers. *EURASIP Journal on Audio, Speech, and Music Processing.* Doi: https://doi.org/10.1155/2010/973954/

Zechner, K., Higgins, D., Xi, X., & Williamson, D. M. (2009). Automatic scoring of non-native spontaneous speech in tests of spoken English. *Speech Communication, 51*(10), 883–895. Doi: https://doi.org/10.1016/j.specom.2009.04.009