Cogent Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/oaed20

A systematic literature review of research on automatic speech recognition in EFL pronunciation

Yao Liu, Faizahani binti Ab Rahman & Farah binti Mohamad Zain

To cite this article: Yao Liu, Faizahani binti Ab Rahman & Farah binti Mohamad Zain (2025) A systematic literature review of research on automatic speech recognition in EFL pronunciation, Cogent Education, 12:1, 2466288, DOI: 10.1080/2331186X.2025.2466288

To link to this article: https://doi.org/10.1080/2331186X.2025.2466288

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 16 Feb 2025.

Submit your article to this journal

Article views: 324

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=oaed20

COGENT EDUCATION
2025, VOL. 12, NO. 1, 2466288
https://doi.org/10.1080/2331186X.2025.2466288

LANGUAGE EDUCATION | REVIEW ARTICLE

OPEN ACCESS Check for updates

A systematic literature review of research on automatic speech recognition in EFL pronunciation

Yao Liu[a,b], Faizahani binti Ab Rahman[a] and Farah binti Mohamad Zain[a]

[a]School of Education, Universiti Utara Malaysia, Sintok, Kedah Darul Aman, Malaysia; [b]School of Foreign Languages, Shandong Women's University, Ji'nan, Shandong, P.R. China

ABSTRACT

To help EFL learners improve English pronunciation skills, researchers have recently turned to automatic speech recognition (ASR) technology. Despite some attempted review studies, relatively few have targeted ASR in an EFL setting. Therefore, the current study aims to delve into various aspects of ASR implementation, including theoretical foundations, research topics, research methods, target features, measurements, and findings, drawing upon published empirical literature. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses guided the review. Of the 6138 articles with 2886 from Web of Science Core Collection and 3252 from Scopus, a total of 24 empirical articles were included. The results revealed: (1) twenty percent of the studies consisted of underpinning theories/models/concepts for the implementation of ASR in an EFL context; (2) a majority of studies adopted a quasi-experimental design, primarily focusing on the pronunciation gains in accuracy; (3) most learning activities with the use of ASR tools were designed for in-class activities; (4) the reviewed ASR studies were primarily concerned with the production accuracy of segmental features. Based on the findings, ASR is proven to be an effective tool in language classrooms but entails limitations. Future studies should investigate how ASR can best be implemented in a classroom setting.

ARTICLE HISTORY
Received 3 April 2024
Revised 21 November 2024
Accepted 7 February 2025

KEYWORDS
Automatic speech recognition; EFL; pronunciation; systematic literature review

SUBJECTS
Communication Technology; Artificial Intelligence; Study of ODL and eLearning; Study Skills; Education - Social Sciences; Language Teaching & Learning

1. Introduction

At present, for many EFL learners, achieving intelligible and comprehensible pronunciation remains a challenging task. It is well-acknowledged that pronunciation learning requires certain phonetic knowledge, extensive practice, and feedback (McCrocklin, 2019a). Traditionally, pronunciation learning has frequently involved practice with a teacher who provides individualized assistance and, where necessary, timely feedback (Ding et al., 2019; Hincks, 2003). According to skill acquisition theory (DeKeyser, 2020), when individuals have insufficient time for deliberate practice and corrective feedback, the proceduralization of 'knowledge that' (phonetic knowledge) into 'knowledge how' (procedural knowledge) becomes challenging, and consequently it's difficult to acquire production behaviors or skills (Zhu et al., 2024). However, in traditional EFL classrooms, giving each student timely, personalized feedback seems time-consuming and unworkable in conventional EFL classes since the class size is large and the in-class time is limited (Georgiou, 2019; Sardegna et al., 2018). Furthermore, EFL learners may not receive enough opportunities for in-class interaction (Chang et al., 2012; Hsu, 2016) and practice (Ahn & Lee, 2016).

Against this backdrop, researchers have shifted their attention to technologies, which can address the aforementioned issues that challenge traditional face-to-face classrooms (Rehman, 2021). One emerging technology is automatic speech recognition (ASR) technology (Lai & Chen, 2022; McCrocklin, 2019a), which has increasingly been recognized for its beneficial features, such as the unlimited practice coupled with immediate and personalized feedback (Dai & Wu, 2023; Evers & Chen, 2022; McCrocklin, 2019a; Ngo et al., 2024). As mobile phones, tablets, and other such devices grow in popularity, ASR-dictation

CONTACT Yao Liu liu_yao@ahsgs.uum.edu.my School of Education, Universiti Utara Malaysia, Sintok, Kedah Darul Aman, Malaysia

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

technology embedded in these devices is increasingly accessible to language learners. It generates immediate and visual feedback by converting speech into text transcripts without offering speech analysis (Inceoglu et al., 2023; Liu et al., 2022). In this way, learners get access to evaluate their pronunciation accuracy by contrasting what they say with the corresponding transcribed text it generates and adjusting accordingly (Lai & Chen, 2022).

2. Literature review

2.1. Research on ASR for pronunciation learning and teaching

Recent years have witnessed an increase in empirical studies that examined ASR usage in the EFL/ESL pronunciation learning contexts (Bashori et al., 2024a). Overall, studies have mostly generated favorable learning outcomes (Farrús, 2023; Ngo et al., 2024), even though the transcription accuracy of ASR is not 100 percent (Bashori et al., 2024a). It was employed alone for autonomous practice (e.g. Bashori et al., 2024b), embedded in other learning platforms (e.g. Hsu, 2024; Wang & Young, 2015), or coupled with other learning activities such as peer feedback (e.g. Dai & Wu, 2023; Evers & Chen, 2022). It facilitates the provision of immediate feedback for the segmental aspect of pronunciation (McCrocklin, 2019a; Ngo et al., 2024) and improves learners' pronunciation accuracy in a short time (Bashori et al., 2024b; Dai & Wu, 2023; Inceoglu et al., 2020; McCrocklin, 2019a). In a study, McCrocklin (2019a) illustrated how *Windows Speech Recognition* improves the accuracy of several problematic segmental sounds for ESL learners in an advanced ESL listening course. In Inceoglu et al.'s (2020) study, 19 intermediate Korean EFL learners participated in a 3-week autonomous mobile-based ASR pronunciation training program in which they practiced pronunciation using a short text independently and at their preferred pace. Two sets of minimal pairs (i.e. /i/-/ɪ/and/ɛ/-/æ/) were the target of training. Analysis of the pre-and post-test data demonstrated an enhanced performance for the/i/-/ɪ/contrast, whereas no significant improvement was observed for the/ɛ/-/æ/contrast. Additionally, ASR has been shown to benefit those who struggle with some psychological issues like shyness and anxiety when learning pronunciation (Bashori et al., 2024a; Dai & Wu, 2023; Guskaroska, 2024). Apart from the above-mentioned benefits, ASR usage has been reported to help foster autonomy (e.g. Dai & Wu, 2023; McCrocklin, 2016). The interview findings from Dai and Wu (2023) showed that EFL learners who employed ASR tended to develop a sense of autonomy. Similar findings were reported in a study by McCrocklin (2016) where ESL students in the HYBRID group were reported spending more time on autonomous pronunciation learning.

Whereas ASR has increasingly been proven beneficial, studies have unveiled its deficiencies. For instance, how language learning theories or models are incorporated as pedagogical underpinnings was insufficiently addressed. A similar limitation was highlighted by Hwang et al. (2022), who observed the lack of guiding frameworks for integrating recognition technology into language learning design. Moreover, in line with a recent meta-analysis by Ngo et al. (2024), its limited impact on suprasegmental pronunciation was identified (Farrús, 2023; Hincks, 2003; Kochem et al., 2022). Ample evidence has highlighted the importance of both segmental and suprasegmental features for achieving a comprehensible and intelligible pronunciation (Levis, 2018; Sardegna, 2022). However, there is insufficient evidence on how ASR usage supports the development of intelligibility and comprehensibility, which are increasingly recognized as the primary goals for EFL pronunciation teaching rather than nativeness (Inceoglu et al., 2023; Levis, 2005; Thomson & Derwing, 2015).

Most ASR programs (e.g. dictation ASR) were originally developed for commercial use instead of pronunciation practice, the technology is still limited in its ability to assist learners in correcting specific errors (Evers & Chen, 2022; Ngo et al., 2024; Pennington & Rogerson-Revell, 2019). Many learners struggle to identify the differences between their pronunciation and the correct one, making self-correction challenging (Garcia et al., 2020). Furthermore, Kochem et al. (2022) pointed out a noticeable disparity between the ASR-equipped software used in studies and the tools accessible to teachers and students in educational settings. This highlights the notion that commercial systems and research may still not be fully integrated with each other (Farrús, 2023). This underscores the necessity for more research on the design and application of ASR for pronunciation teaching and learning.

2.2. Recent review studies on ASR for pronunciation

The pedagogical significance of ASR in assisting pronunciation instruction and learning has prompted systematic reviews and meta-analysis (Farrús, 2023; Ngo et al., 2024; Nickolai et al., 2024; Shadiev & Liu, 2023). These studies consistently highlighted its widespread use in pronunciation teaching and learning, particularly for English; review studies also reported positive outcomes, such as increased affective gains (see Nickolai et al., 2024; Shadiev & Liu, 2023).

However, recent reviews (see Figure 1) have varied widely in focus, with limited attention to its application in EFL settings despite its potential to overcome EFL classroom limitations. While two studies explored ASR for general language learning, few specially targeted pronunciation (e.g. Hwang et al., 2022; Shadiev & Liu, 2023). Additionally, the databases included differed, ranging from CALL-specific journals to sources like IEEE Xplore and Google Scholar. This inconsistency limits comparability across studies. Moreover, three studies, such as Farrús (2023), Nickolai et al. (2024), and Ngo et al. (2024), generalized findings to language learners broadly, reducing relevance for EFL contexts, where EFL learners have fewer opportunities to use English outside the classroom than ESL learners. Moreover, recent review studies revealed that few have examined the theoretical frameworks underpinning ASR research, with most focusing primarily on its practical effects. Therefore, to address these gaps, this study offers a targeted synthesis of ASR studies conducted in EFL contexts through a systematic literature review (SLR), guiding future research in this domain. Accordingly, the following research questions were explored:

RQ1: What theoretical frameworks have been applied to ASR studies?

RQ2: What research topics have been addressed?

RQ3: What research methods have been commonly adopted in ASR studies exploring its efficacy on EFL learners' pronunciation?

RQ4: What are the contexts in which ASR is applied for pronunciation practice, and what is the duration of its application?

RQ5: What are targeted phonetic features, and how is learner pronunciation measured in these studies?

3. Methodology

The current study utilized a systematic literature review method. The design process of the systematic literature review undertaken was guided by the PRISMA guidelines (Moher et al., 2010). PRISMA is a globally recognized and validated method that is utilized to perform systematic reviews (Shortt et al., 2023). The guidelines have been successfully implemented in various disciplines, including language learning (e.g. Kernagaran & Abdullah, 2022; Shortt et al., 2023). The aim of using the PRISMA guidelines was to assist the authors in providing a transparent account of the reasoning behind their review and the procedures employed to achieve their research findings (Marco-Lajara et al., 2023). In the study, we followed the four essential stages detailed in the PRISMA procedure. These stages include conducting a comprehensive search

Author(s)	Research Area	Articles Reviewed	Reviewed period	Databases Used
Ngo et al. (2024)	ESL/EFL pronunciation	15	2008-2021	ProQuest, ERIC, Google Scholar, and some SSCI journals
Nickolai et al. (2024)	L2 pronunciation	50	2012-March 2023	Major CALL-specific journals
Farrús (2023)	ESL pronunciation	39	2013-July 2023	Springer Link ; Science Direct; IEEExplore; Google Scholar
Shadiev and Liu (2023)	Language learning	26	2014-2020	WoS
Hwang et al. (2022)	Language learning	48	2011-2020	WoS

Figure 1. Recent reviews on ASR in English language learning.

and identification process, systematically evaluating and extracting relevant information, defining eligibility criteria, and performing quality appraisal and data analysis (Moher et al., 2010).

3.1. Criteria for eligibility

In order to identify the resources as comprehensively as possible, the eligibility of literature for inclusion or exclusion was formulated in terms of document types, the language of related publications, the timespan, and the literature focus. First, peer-reviewed empirical studies utilizing quantitative, qualitative, or mixed methods were included. This selection was prompted by the need to address the inconsistent findings in empirical evidence related to ASR. Examples like unpublished papers, conference papers, proceedings, book reviews/chapters, theses, reports, and other similar materials were excluded. Additionally, review articles (e.g., scoping reviews, meta-analyses, narrative reviews) were excluded as they did not qualify as empirical studies (Han & Gao, 2023). Peer-reviewed articles without full text were also excluded. Second, concerning language, only articles written in English were considered for inclusion, while those published in other languages were excluded. Third, the timespan was set up from the coverage date of the two databases to 2nd November 2024. Fourth, the focus of this review centered on the utilization of ASR for learners' pronunciation in an EFL context. Consequently, articles that related to other languages (e.g. Mandarin, French, Dutch, and Spanish), those focusing on non-pronunciation aspects, or merely citing ASR as an example without empirical analysis were excluded to ensure a relevant scope for this review.

3.2. Literature search and identification

The literature search process took place on 2nd November 2024. All searches were conducted based on title, keywords, and abstract. First, we searched articles published in two major electronic databases (i.e. Web of Science and Scopus) from their inception up to 2nd November 2024. These two databases, extensively recognized for their utility in academic literature searches and selection, are frequently utilized in systematic reviews across a variety of academic disciplines, including language education (e.g. Barrot, 2022; Demir & Kartal, 2022). Other databases, such as Google Scholar, were excluded due to inconsistent accuracy and inadequate provision of comprehensive citation data (Farooq, 2023). Consequently, we believed the inclusion of these two databases was both reasonable and comprehensive.

The retrieval formulae were formulated using Boolean expressions (see Figure 2) within the advanced search mode. Building upon previous studies concerning ASR and pronunciation (e.g. Ngo et al., 2024), the combination of ASR-related and pronunciation-related words was used as keywords for the searches. The same search string was employed in both databases. This approach was employed to refine and customize the search criteria. More precisely, in the Web of Science Database, the research was limited to all articles. Subsequently, the results were refined by limiting the research to 'Web of Science Core Collection' under the 'Database' option, filtering articles written in English through the 'Languages' option. In the Scopus database, the search was restricted to 'article' document type, and articles in the English language. The Boolean terms generated 6138 results in total, with 3252 results in Scopus and 2886 in Web of Science Core Collection.

3.3. Screening and extraction

The overall process was guided by the recommendations put forth by Kitchenham et al. (2009). Specifically, the data extraction was performed by one author, LIU YAO, with verification of the

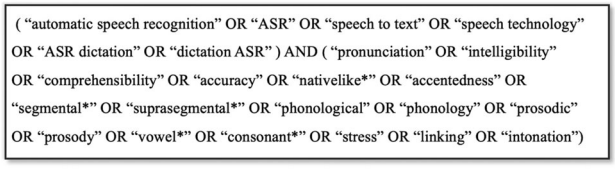


Figure 2. Boolean expressions employed in the study.

extraction accuracy performed by research assistant LI QI. In instances of disagreement, discussions had been carried out until a consensus was achieved and agreed upon (Iden & Eikebrokk, 2013). The retrieved information was input into Endnote 21 and underwent three rounds of rigorous screening and extraction procedures.

Initially, upon eliminating duplicate cases (which brought the total records down from 6138 to 3863), we proceeded to review the article titles and abstracts to determine whether they were relevant. Subsequently, the collected records, as depicted in Figure 3, were evaluated using the inclusion/exclusion criteria based on their titles and abstracts. If it was obvious that an article failed to satisfy the inclusion criteria, it was removed. Ultimately, a total number of 54 articles passed these criteria, with 24 deemed eligible for inclusion in the review.

3.4. Quality appraisal and data analysis

To evaluate the quality of the reviewed articles, we used the Mixed Method Appraisal Tool (version 2018) developed by Hong et al. (2018). This tool has been utilized in previous systematic reviews to evaluate the quality of empirical studies (see Fisher et al., 2024; Han & Gao, 2023). There are three main steps involved in employing the MMAT: responding to two screening questions; determining the appropriate research study category (qualitative, quantitative, or mixed methods), and evaluating the quality of studies through response options ('Yes' if they fulfill the criteria, 'No' if they do not, 'Can't tell' if it is unclear from the reporting whether the criteria are met. Following an extensive examination, the findings showed that 24 articles were identified as satisfying the inclusion criteria and were subsequently incorporated in the final sample of this review.

Following the quality appraisal, we analyzed the 24 articles using content analysis. The first researcher and the research assistant independently coded the parts related to the theoretical framework, research topics, research methods, ASR application, treatment duration, targeted pronunciation features, and pronunciation measurement. The final coding results were compared and discussed. Instances of any differences were resolved via consulting the other two researchers.

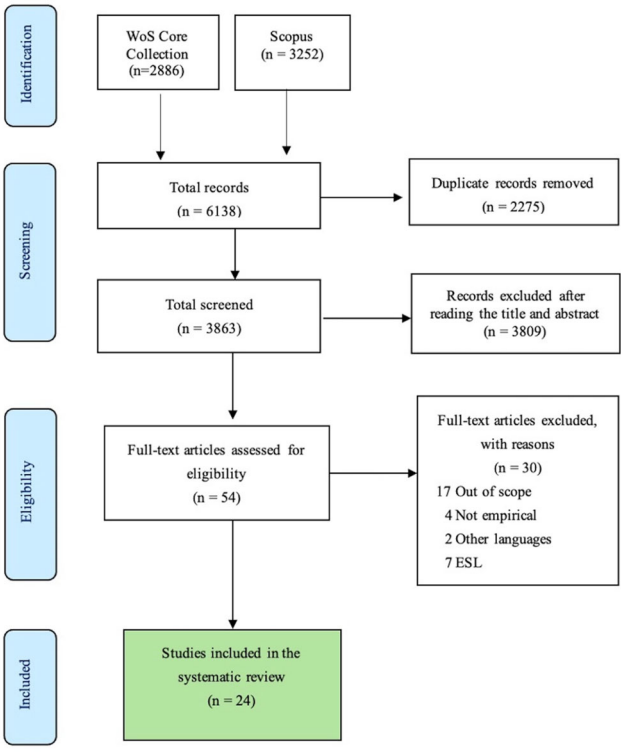


Figure 3. PRISMA flow diagram.

4. Results

All the 24 retrieved articles (marked with an * in the references) were primary studies that specifically examined the application of ASR in an EFL context, with publication dates ranging from 2003 to 2024.

4.1. Theoretical framework underpinning ASR studies

As presented in Table 1, most of the studies (17 out of 24) did not incorporate relevant theories, models, or concepts into their research. Only six studies (n = 6) explicitly mentioned relevant theories. Those theories were adopted to explain the impact of a single feedback mode (i.e. ASR alone) or a combination mode (i.e. ASR and peer feedback) on EFL learners' pronunciation performance, or to explain their perceptions of ASR acceptance. For instance, Dai and Wu (2023) drew on skill acquisition theory (SAT) to explain the benefits of ASR, highlighting its capacity to offer abundant practice opportunities for pronunciation without time and space constraints. Two other studies adopted socio-cultural theory (SCT) to explore how the combination mode contributes to pronunciation learning. These studies suggested that learners could enhance their pronunciation performance through meaningful interactions with more capable peers. In addressing the design of necessary feedback in ASR-based CALL systems, one study introduced the concept of explicit feedback and evaluated its efficacy in pronunciation learning. In another study, guided by audio-lingual theory, Hincks (2003) assessed the effectiveness of *Talk to Me*, an ASR software, on learners' pronunciation following a five-hour training session. Unlike previous studies, two studies drew on the technology acceptance model (TAM) without examining its effects on actual pronunciation learning outcomes. Hsu (2024) explored factors influencing EFL learners' acceptance of ASR usage while Hsu (2016) investigated the structural relationship between TAM and the perceptual learning styles in EFL students.

4.2. Research topics in ASR studies

The topics of the 24 studies were grouped into six main categories. As shown in Figure 4, 67% (n = 16) of the total studies focused on the pedagogical efficacy of ASR in relation to EFL learners' pronunciation performance. The evaluation of the accuracy of ASR technology was the focus of two studies (n = 2), which reflected recent research attention in this area. An investigation conducted by Lai and Chen (2022) examined the recognition accuracy of three popular ASR tools. They aimed to evaluate whether these tools could accurately transcribe proficient young EFL learners' spoken output and determine if there were differences in recognition accuracy between learners with various proficiency levels. In addition, there are five studies (n = 5) investigating the relationship between rater ratings and ASR-scored pronunciation. For example, Kim (2006) found that the scores assigned by a specific ASR technology (i.e. FluSpeak) correlated with the scores assigned by native speakers. Based on the findings, Kim (2006) suggested that FluSpeak might be an effective tool for instructing EFL students in pronunciation when instructors who are native English speakers are not easily accessible. Out of the 24 reviewed articles, three studies (n = 3) focused on examining EFL learners' acceptance of ASR technology (Dillon & Wells, 2021; Hsu, 2016, 2024). Besides, it is interesting to highlight that a small proportion of studies examined more than one research topic. For instance, Inceoglu et al. (2020) investigated how ASR technology impacted learner' beliefs and their pronunciation performance. In another study, Hsu (2016) delved into exploring whether EFL learners' perceptual styles affected the utilization of ASR technology.

Table 1. Theories/models/concepts used.

Theories/models/concepts	Studies
SCT	Sun (2023); Dai and Wu (2023)
SAT	Dai and Wu (2023)
Explicit feedback	Wang and Young (2015)
Audio-lingual theory	Hincks (2003)
TAM	Hsu (2024); Hsu (2016)

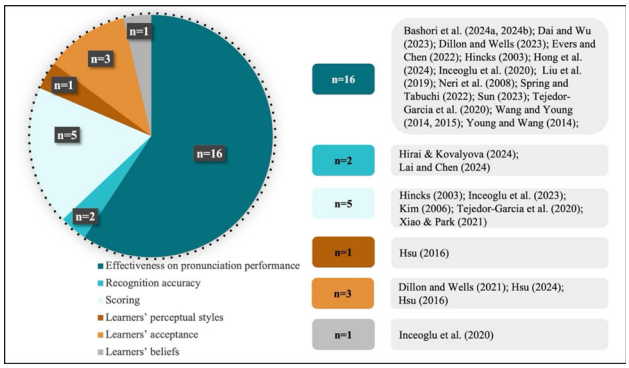


Figure 4. Research topics in ASR studies.

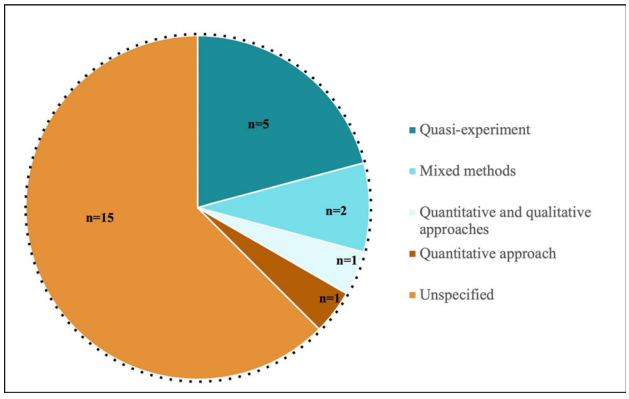


Figure 5. Research methods used in ASR studies.

4.3. Research methods employed in ASR studies

Nine out of the 24 studies mentioned exactly what research method they employed in the studies (see Figure 5). Two main research methods, quasi-experimental and mixed methods, were identified. However, many studies (n = 15) did not explicitly specify the research methods employed, often only describing data collection techniques such as questionaries (e.g. Dillon & Wells, 2021; Hsu, 2016, 2024) and pre-/posttest (e.g. Bashori et al., 2024a; Liu et al., 2019; Neri et al., 2008).

4.4. Application and treatment duration in ASR studies

The findings demonstrated the diverse application of ASR both within and beyond traditional classroom settings. While ASR was primarily utilized for autonomous learning and practice outside class (n = 16), notable instances (n = 8) arose where ASR was employed during in-class time. As to treatment duration, three studies (see Hirai & Kovalyova, 2024; Kim, 2006; Lai & Chen, 2022) were excluded from this analysis because they did not report treatment durations. An analysis of the reviewed studies (21 out of 24) revealed that treatment durations varied significantly, ranging from one week to 15 weeks (see Figure 6). Following Ngo et al. (2024), we categorized the length of treatment into short (1 to 4 weeks), medium (5 to 8 weeks), and long (9 weeks and more). Studies conducted by Hincks (2003) and Hong et al. (2024) were classified under the short-duration category, given that EFL students typically received one and a half hours of in-class sessions per week (Figure 6).

4.5. Target features and measurements in ASR studies

The targeted features were coded on three levels (see Table 2). An analysis revealed that eleven studies (n = 11) examined segmental features, specifically minimal pairs or commonly mispronounced phonemes, elicited primarily from controlled tasks classified by Saito and Plonsky (2019). We also found that 13 studies did not mention the phonetic features they targeted. For example, in a study by Wang and

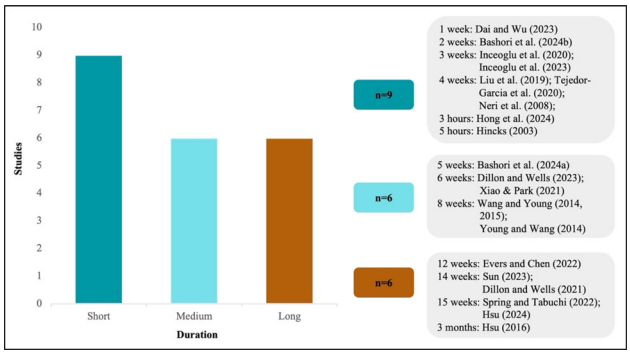


Figure 6. Duration of the treatment in ASR studies.

Table 2. Targeted features in ASR studies.

	Classification	Studies	*N*
Targeted features	Segmental features	Bashori et al. (2024b); Dai and Wu (2023); Dillon and Wells (2023); Evers and Chen (2022); Inceoglu et al. (2020);Liu et al. (2019); Neri et al. (2008); Spring and Tabuchi (2022); Sun (2023); Tejedor-Garcia et al. (2020); Xiao and Park (2021)	11
	Suprasegmental features	Dai and Wu (2023)	1
	Not mentioned	Bashori et al. (2024a); Dillon and Wells (2021); Hincks (2003); Hirai and Kovalyova (2024); Hong et al. (2024); Hsu (2016); Hsu (2024); Inceoglu et al. (2023); Kim (2006); Lai and Chen (2022); Wang and Young (2014, 2015); Young and Wang (2014)	13

Table 3. Measurements of pronunciation in ASR studies.

Measurements		Example studies
Accuracy		
	Word-level accuracy	Bashori et al. (2024b); Wang and Young (2014)
	Segmental accuracy	Dai and Wu (2023); Dillon and Wells (2023);
	Word stress accuracy	Dai and Wu (2023)
Comprehensibility		Sun (2023); Bashori et al. (2024a)
Accentedness		Evers and Chen (2022); Bashori et al. (2024a)
Intelligibility		Inceoglu et al. (2023)
Pronunciation quality of words		Neri et al. (2008)

Young (2015), pronunciation accuracy was measured based on eight sentences without specifying the targeted phonetic features.

In terms of pronunciation measurements, the outcome measures varied greatly depending on the types of features targeted, which were coded into six levels (see Table 3). Pronunciation accuracy was further divided into segmental accuracy, word-level accuracy, and word stress accuracy. In this section, three studies (Dillon & Wells, 2021; Hsu, 2016, 2024) were excluded due to the absence of specific measurements of learner pronunciation.

5. Discussion

Examining the 24 empirical studies revealed a consistent growth in ASR research for EFL pronunciation learning and teaching. Notably, the period from 2021 to 2024 saw a particularly productive phase, with 14 out of the 24 articles published during this time interval. This further highlights the increasing interest and improvements in this domain of research.

5.1. Results of the review

Our findings identified SAT and TAM as the primary theoretical underpinnings of ASR research. These theories explicitly explained the positive effects of integrating ASR within and beyond classrooms. For instance, SAT underscores the pivotal role of feedback and practice (Suzuki, 2023; Zhu et al., 2024).

ASR provides ample opportunities for learners to repeatedly practice and internalize their pronunciation, enabling them to develop procedural skills and automaticity (Dai & Wu, 2023; Zhu et al., 2024). Such highlights the compatibility between SAT and the notion that focused, intentional practice is essential for developing learners' pronunciation using ASR technology. Additionally, scholars (Hsu, 2016, 2024) examined ASR acceptance using the TAM, which provides a framework for understanding why learners use it (Guskaroska, 2024). While previous studies often relied on qualitative methods (e.g. interviews) to capture learners' perceptions (see Dai & Wu, 2023; Evers & Chen, 2022; McCrocklin, 2019b), they lacked a robust theoretical framework to explore this process. The adoption of TAM would help explore how its core concepts influence ASR acceptance (Guskaroska, 2024).

Concerning the second research question, we found that most research on the effectiveness of ASR for training at both the segmental and word levels has been considered (e.g. Inceoglu et al., 2020; Tejedor-Garcia et al., 2020). Meanwhile, researchers have recently grown interest in evaluating ASR transcription accuracy (e.g. Lai & Chen, 2022) and comparing it with human raters, raising the potential for its application in formative assessments. This interest stemmed from the recognition that most ASR programs were initially designed for commercial use, leading to limitations in the accuracy of feedback provided to learners (Dillon & Wells, 2021; McCrocklin, 2019a). Considering only practicing with ASR would not be that feasible for every learner as some learners still require assistance to correct pronunciation mistakes (Evers & Chen, 2022), several researchers (i.e. Dai & Wu, 2023; Evers & Chen, 2022; Sun, 2023) contended that incorporating peer feedback alongside dictation ASR could enhance its effectiveness. Whereas studies have revealed the merits of this mode, they are limited in number, and more research in this regard is required.

Regarding the third research question, quasi-experimental designs with pre-tests and post-tests were the most frequently employed research designs in studies. The findings were comparable to those reported by Amini Farsani and Babaii and Vitta et al. (2022) since instructors cannot randomly assign individuals. Additionally, mixed-methods design started to draw researchers' attention (see Dai & Wu, 2023; Sun, 2023) as it can avoid a partial understanding of the studies with the use of a single approach (Creswell & Creswell, 2023).

In terms of research question four, findings revealed that ASR technology was primarily employed for self-learning outside the classroom with treatment durations varying significantly based on instructional focus and treatment contexts. Utilizing ASR outside class allows learners to self-regulate their learning at their own pace while freeing up class time for other activities (McCrocklin, 2016; 2019a). However, whether ASR can be considered a substitute for classroom instruction requires further research.

As to the fifth research question, most ASR studies primarily focused on accuracy in producing segmental features. Surprisingly, some recent studies have started employing comprehensibility and accentedness as additional measures of pronunciation performance (see Dai & Wu, 2023; Sun, 2023). With English speaking skills gaining greater importance in both professional and academic settings, issues of pronunciation intelligibility and comprehensibility are becoming more prominent (Newton, 2017). Given the ultimate goal of communication is to be understood (Jenkins, 2000; Levis, 2005), our findings suggest that future pronunciation research on ASR should assess learners' pronunciation in terms of comprehensibility and intelligibility.

5.2. Recommendations for future research

The results of the review identified several areas that deserved further exploration.

5.2.1. Studies with longer treatments

Pronunciation learning needs a longer duration (Hong et al., 2024). The variability in treatment duration observed in the reviewed studies, spanning from a short duration (e.g. Bashori et al., 2024b; Dai & Wu, 2023) to longer periods (e.g. Evers & Chen, 2022; Sun, 2023), highlights a potential limitation in designing effective ASR-based pronunciation activities in real-classroom contexts. Classroom studies are deemed more replicable than those conducted in labs and may provide practical value for instructors to teach pronunciation systematically (Saito & Lyster, 2012; Thomson & Derwing, 2015). To ensure replicability in real-school contexts, the intervention should last at least 12 weeks as short-duration studies

typically yield significant effects due to novelty or high fidelity (Ni et al., 2023). Therefore, longer, classroom-based studies are recommended.

5.2.2. Studies with diverse pronunciation tasks

Whereas students may perform well in controlled tasks (such as reading-aloud isolated words) and show improved pronunciation with the help of ASR, this success doesn't necessarily imply that they can equally perform well in spontaneous tasks. In spontaneous connected speech, additional factors like co-articulation (the influence of one sound on the articulation of another) and sentence accent come into play, which may present different challenges compared to controlled tasks. This underscores the significance of considering a range of speech tasks when assessing pronunciation skills. The use of a reading-aloud task and a spontaneous task can provide a comprehensive evaluation of learners' pronunciation (Evers & Chen, 2022; Saito & Plonsky, 2019). Reading-aloud tasks function as a controlled, form-focused activity designed to elicit awareness and analysis of pronunciation forms (Lee et al., 2020). In contrast, spontaneous tasks shift the focus from form (i.e. pronunciation features) to meaning (Celce-Murcia et al., 2010), allowing pronunciation to be measured in a relatively natural, unconscious manner (Hu et al., 2022). This is a way to test whether learners can shift their gains from form to meaning (Dalman, 2022; Gooch et al., 2016), as demonstrated in previous studies (e.g. Evers & Chen, 2022; Sun, 2023).

5.2.3. Research on ASR in flipped pronunciation classrooms

In pronunciation learning and teaching, immediate feedback is particularly crucial for self-practice (McCrocklin, 2019b; Ngo et al., 2024; Saito & Lyster, 2012). Immediate feedback can alert EFL students on their pronunciation problems, which is the first necessary step for them to remedy those problems (Bodnar et al., 2017; Neri et al., 2008). ASR can support this process by enabling repeated practice and helping learners identify areas for improvement (Franco et al., 2010; Jiang et al., 2023b). Despite researchers highlighting the role of ASR technology in pronunciation learning, it faces two main challenges in an actual classroom. First, pronunciation instruction is mostly integrated into other courses and the implementation of ASR technology during in-class time faces time constraints (McCrocklin, 2019a). Second, ASR mainly deals with lower cognitive skills based on Bloom's revised taxonomy (Jiang et al., 2022; 2023b). In that case, more in-class time should be freed up for knowledge application.

As highlighted by Jiang et al. (2023b), ASR's benefits can be maximized in a flipped classroom (FC), because learners in a FC are expected to spend some time conducting pre-class activities designed to cover lower cognitive skills prior to attending class. More in-class time can be freed up for higher cognitive skills based on Bloom's revised taxonomy. Its integration into the pre-class phase has shown positive effects on learners' pre-class engagement and better performance in vocabulary, oral speaking, and willingness to communicate (see Jiang et al., 2022; 2023a; Jiang et al., 2023b). These findings indicate that ASR may be also applicable in a flipped pronunciation classroom, which deserves further investigation.

5.2.4. Studies on the ASR-based AI applications

Despite the advancements in ASR technology's reliability (Hasan et al., 2017), recent studies reveal a persistent gap in its effectiveness at prosodical features (Farrús, 2023; Kochem et al., 2022; O'Brien et al., 2022). Besides, ASR is unsuitable for automatically evaluating learner input (Bogach et al., 2021). As AI-enabled applications can better handle prosodic features and then respond accordingly (Chung & Bong, 2022), advances in AI technologies have enhanced ASR's ability to assist pronunciation learning and teaching by addressing its limitations and enabling more personalized learning experiences. Hong et al. (2024), for instance, examined how educational robots, developed using Google's natural language and speech-to-text technology, improved EFL learners' pronunciation accuracy after two rounds of self-learning. Fu et al. (2020) explored the feasibility of applying an AI-enabled automatic scoring app that relies upon ASR and automatic pronunciation evaluation techniques to support pronunciation learning. Even though there are separate studies on ASR and AI-enabled applications for pronunciation learning, less evidence has shown on the usage of ASR-based AI-enabled application. Future research may look into the role of ASR-based AI applications and their effects on pronunciation learning outcomes.

6. Conclusion and limitations

The purpose of this systematic review is to provide (1) a comprehensive summary of ASR usage for EFL pronunciation; and (2) a summary of research gaps and future directions for further improvement. Overall, ASR has increasingly been recognized for improving the benefits that learners and teachers value. Through addressing the research questions, some conclusions and research gaps have emerged: (1) the use of ASR for improving pronunciation has attracted researchers' attention in the EFL context; (2) whereas the measurements of pronunciation performance were different, the effectiveness of this learning strategy was observed; (3) notable research gap exists in the limited incorporation of theories, models, or concepts into studies exploring its effectiveness in pronunciation learning; (4) while most studies were conducted to explore learners' pronunciation performance and their perceptions, there has been a notable absence of attention to teachers' perspective; (5) studies indicated that employing a combination feedback mode (e.g. peer plus ASR feedback) could contribute to more pronunciation gains compare to using a single feedback mode; (6) ASR has primarily been integrated into traditional language classrooms. To optimize the benefits of ASR, integrating it into the pre-class phase of a flipped pronunciation classroom was proposed.

With findings reached and research gaps found, several research limitations have been acknowledged. On the one hand, the scope of this review is limited to articles and does not include a larger range of databases. Follow-up reviews can broaden the scope to include more databases, as well as more types of articles, such as conference papers and theses, to determine whether the findings are consistent over a broader range of studies. On the other hand, as this review exclusively incorporated articles written in English, studies published in other languages, such as Mandarin, might have been neglected. For the authors in the study, future studies may explore relevant studies written in their own languages to unveil a complete picture of ASR studies.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Shandong Women's University, China; Shandong Women's University, China.

About the authors

Yao Liu is a PhD candidate at the School of Education, Universiti Utara Malaysia. Additionally, she serves as a lecturer in the School of Foreign Languages at Shandong Women's University, China. Her research interests include pronunciation instruction and Mobile-Assisted Language Learning (MALL).

Faizahani binti Ab Rahman is a senior lecture at the School of Education, Universiti Utara Malaysia. Dr Faizahani holds a PhD from University of Reading, United Kingdom. Her research areas include TESOL and gender studies and issues.

Farah binti Mohamad Zain is an associate professor at the School of Education, Universiti Utara Malaysia. Dr Farah holds a PhD from Universiti Sains Malaysia. Her research areas include digital content development, instructional technology, and augmented reality.

References

*Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024a). I Can Speak: Improving English pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching*, *18*(5), 443–461. https://doi.org/10.1080/17501229.2024.2315101

*Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024b). 'Look, I can speak correctly': Learning vocabulary and pronunciation through websites equipped with automatic speech recognition technology. *Computer Assisted Language Learning*, *37*(5–6), 1335–1363. https://doi.org/10.1080/09588221.2022.2080230

*Dai, Y., & Wu, Z. (2023). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning*, *36*(5–6), 861–884. https://doi.org/10.1080/09588221.2021.1952272

*Dillon, T., & Wells, D. (2021). Student perceptions of mobile automated speech recognition for pronunciation study and testing. *English Teaching*, *76*(4), 101–122. https://doi.org/10.15858/engtea.76.4.202112.101

*Dillon, T., & Wells, D. (2023). Effects of pronunciation training using automatic speech recognition on pronunciation accuracy of Korean English language learners. *English Teaching*, *78*(1), 3–23. https://doi.org/10.15858/engtea.78.1.202303.3

*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*(8), 1869–1889. https://doi.org/10.1080/09588221.2020.1839504

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

*Hirai, A., & Kovalyova, A. (2024). Speech-to-text applications' accuracy in English language learners' speech transcription. *Language Learning & Technology*, *28*(1), 1–21.

*Hong, Z.-W., Tsai, M.-H. M., Ku, C. S., Cheng, W. K., Chen, J.-T., & Lin, J.-M. (2024). Utilizing robot-tutoring approach in oral reading to improve Taiwanese EFL students' English pronunciation. *Cogent Education*, *11*(1).1–15. https://doi.org/10.1080/2331186X.2024.2342660

*Hsu, H.-W. (2024). An examination of automatic speech recognition (ASR)-based computer-assisted pronunciation training (CAPT) for less-proficient EFL students using the technology acceptance model. *International Journal of Technology in Education*, *7*(3), 456–473. https://doi.org/10.46328/ijte.681

*Hsu, L. (2016). An empirical examination of EFL learners' perceptual learning styles and acceptance of ASR-based computer-assisted pronunciation training. *Computer Assisted Language Learning*, *29*(5), 881–900. https://doi.org/10.1080/09588221.2015.1069747

*Inceoglu, S., Chen, W. H., & Lim, H. (2023). Assessment of L2 intelligibility: Comparing L1 listeners and automatic speech recognition. *ReCALL*, *35*(1), 89–104. https://doi.org/10.1017/S0958344022000192

*Inceoglu, S., Lim, H., & Chen, W. H. (2020). ASR for EFL pronunciation practice: Segmental development and learners' beliefs. *The Journal of AsiaTEFL*, *17*(3), 824–840. https://doi.org/10.18823/asiatefl.2020.17.3.5.824

*Kim, I. S. (2006). Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation. *Educational Technology & Society*, *9*(1), 322–334.

*Lai, K.-W. K., & Chen, H.-J. H. (2022). An exploratory study on the accuracy of three speech recognition software programs for young Taiwanese EFL learners. *Interactive Learning Environments*, *32*(5), 1–15. https://doi.org/10.1080/10494820.2022.2122511

*Liu, X., Xu, M., Li, M., Han, M., Chen, Z., Mo, Y., Chen, X., & Liu, M. (2019). Improving English pronunciation via automatic speech recognition technology. *International Journal of Innovation and Learning*, *25*(2), 126–140. https://doi.org/10.1504/IJIL.2019.097674

*Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, *21*(5), 393–408. https://doi.org/10.1080/09588220802447651

*Spring, R., & Tabuchi, R. (2022). The role of ASR training in EFL pronunciation improvement: An In-depth look at the impact of treatment length and guided practice on specific pronunciation points. *Computer Assisted Language Learning Electronic Journal*, *23*(3), 163–185.

*Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. *Frontiers in Psychology*, *14*, 1210187. https://doi.org/10.3389/fpsyg.2023.1210187

*Tejedor-Garcia, C., Escudero-Mancebo, D., Camara-Arenas, E., Gonzalez-Ferreras, C., & Cardenoso-Payo, V. (2020). Assessing pronunciation improvement in students of English using a controlled computer-assisted pronunciation tool. *IEEE Transactions on Learning Technologies*, *13*(2), 269–282. https://doi.org/10.1109/TLT.2020.2980261

*Wang, Y.-H., & Young, S. S.-C. (2014). A study of the design and Implementation of the ASR-based iCASL system with corrective feedback to facilitate English learning. *Educational Technology & Society*, *17*(2), 219–233.

*Wang, Y.-H., & Young, S. S.-C. (2015). Effectiveness of feedback for enhancing English pronunciation in an ASR-based CALL system. *Journal of Computer Assisted Learning*, *31*(6), 493–504. https://doi.org/10.1111/jcal.12079

*Xiao, W., & Park, M. (2021). Using automatic speech recognition to facilitate English pronunciation assessment and learning in an EFL context: Pronunciation error diagnosis and pedagogical implications. *International Journal of Computer-Assisted Language Learning and Teaching*, *11*(3), 74–91. https://doi.org/10.4018/IJCALLT.2021070105

*Young, S. S.-C., & Wang, Y.-H. (2014). The game embedded CALL system to facilitate English vocabulary acquisition and pronunciation. *Educational Technology & Society*, *17*(3), 239–251.

Ahn, T. y., & Lee, S.-M. (2016). 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

Barrot, J. S. (2022). Social media as a language learning environment: A systematic review of the literature (2008-2019). *Computer Assisted Language Learning*, *35*(9), 2534–2562. https://doi.org/10.1080/09588221.2021.1883673

Bodnar, S., Cucchiarini, C., Penning de Vries, B., Strik, H., & van Hout, R. (2017). Learner affect in computerised L2 oral grammar practice with corrective feedback. *Computer Assisted Language Learning*, *30*(3–4), 223–246. https://doi.org/10.1080/09588221.2017.1302964

Bogach, N., Boitsova, E., Chernonog, S., Lamtev, A., Lesnichaya, M., Lezhenin, I., Novopashenny, A., Svechnikov, R., Tsikach, D., Vasiliev, K., Pyshkin, E., & Blake, J. (2021). Speech processing for language learning: A practical approach to computer-assisted pronunciation teaching. *Electronics*, *10*(3), 235. https://doi.org/10.3390/electronics10030235

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

Chang, C.-C., Yan, C.-F., & Tseng, J.-S. (2012). Perceived convenience in an extended technology acceptance model: Mobile technology and English learning for college students. *Australasian Journal of Educational Technology*, *28*(5), 809–826. https://doi.org/10.14742/ajet.818

Chung, B., & Bong, H. K. M. (2022). A study on the intelligibility of Korean-accented English: Possibilities of implementing AI applications in English education. *The Journal of AsiaTEFL*, *19*(1), 197–215. https://doi.org/10.18823/asiatefl.2022.19.1.12.197

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). Sage Publications.

Dalman, M. (2022). *Online-based intelligibility instruction for second language (L2) learners (Publication Number 29207245)* [Doctoral thesis]. Northern Arizona University]. ProQuest Dissertations and Theses Global.

DeKeyser, R. (2020). Skill acquisition theory. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition* (3rd ed., pp. 83–104). Routledge.

Demir, Y., & Kartal, G. (2022). Mapping research on L2 pronunciation: A bibliometric analysis. *Studies in Second Language Acquisition*, *44*(5), 1211–1239. https://doi.org/10.1017/S0272263121000966

Ding, S., Liberatore, C., Sonsaat, S., Lučić, I., Silpachai, A., Zhao, G., Chukharev-Hudilainen, E., Levis, J., & Gutierrez-Osuna, R. (2019). Golden speaker builder–An interactive tool for pronunciation training. *Speech Communication*, *115*, 51–66. https://doi.org/10.1016/j.specom.2019.10.005

Farooq, R. (2023). Knowledge management and performance: A bibliometric analysis based on Scopus and WOS data (1988–2021). *Journal of Knowledge Management*, *27*(7), 1948–1991. https://doi.org/10.1108/JKM-06-2022-0443

Farrús, M. (2023). Automatic speech recognition in L2 learning: A review based on PRISMA methodology. *Languages*, *8*(4), 242. https://doi.org/10.3390/languages8040242

Fisher, R., Tran, Q., & Verezub, E. (2024). Teaching English as a foreign language in higher education using flipped learning/flipped classrooms: A literature review. *Innovation in Language Learning and Teaching*, *18*(4), 332–351. https://doi.org/10.1080/17501229.2024.2302984

Franco, H., Bratt, H., Rossier, R., Rao Gadde, V., Shriberg, E., Abrash, V., & Precoda, K. (2010). EduSpeak®: A speech recognition and pronunciation scoring toolkit for computer-aided language learning applications. *Language Testing*, *27*(3), 401–418. https://doi.org/10.1177/0265532210364408

Fu, S., Gu, H., & Yang, B. (2020). The affordances of AI-enabled automatic scoring applications on learners' continuous learning intention: An empirical study in China. *British Journal of Educational Technology*, *51*(5), 1674–1692. https://doi.org/10.1111/bjet.12995

Garcia, C., Nickolai, D., & Jones, L. (2020). Traditional versus ASR-based pronunciation instruction: An empirical study. *CALICO Journal*, *37*(3), 213–232. https://doi.org/10.1558/cj.40379

Georgiou, G. P. (2019). EFL teachers' cognitions about pronunciation in Cyprus. *Journal of Multilingual and Multicultural Development*, *40*(6), 538–550. https://doi.org/10.1080/01434632.2018.1539090

Gooch, R., Saito, K., & Lyster, R. (2016). Effects of recasts and prompts on L2 pronunciation development: Teaching English/ɹ/to Korean adult EFL learners. *System*, *60*, 117–127. https://doi.org/10.1016/j.system.2016.06.007

Guskaroska, A. (2024). *Exploring technology acceptance of ASR for pronunciation learning (Publication Number 31332589)* [Doctoral dissertation]. Iowa State University. ProQuest Dissertations and Theses Global.

Han, J., & Gao, C. (2023). Teachers' achievement goal orientations: A systematic review of 15 years of published empirical research. *Teaching and Teacher Education*, *128*, 104146. https://doi.org/10.1016/j.tate.2023.104146

Hasan, R., Hussein, H., Lazaridis, P., Khwandah, S., Ritter, M., & Eibl, M. (2017 *Improvement of speech recognition results by a combination of systems* [Paper presentation]. 2017 23rd International Conference on Automation and Computing (ICAC). https://doi.org/10.23919/IConAC.2017.8082082

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., & Nicolau, B. (2018). *Mixed methods appraisal tool (MMAT), version 2018. Registration of Copyright (# 1148552)*. Canadian Intellectual Property Office, Industry Canada.

Hu, X., Kuo, L.-J., & Dixon, L. Q. (2022). Effectiveness of interventions on improving English language learners' intelligibility and comprehensibility: A meta-analytic review. *Language Teaching Research*, 1–25. https://doi.org/10.1177/13621688221106832

Hwang, W.-Y., Nguyen, V.-G., & Purba, S. W. D. (2022). Systematic survey of anything-to-text recognition and constructing its framework in language learning. *Education and Information Technologies*, *27*(9), 12273–12299. https://doi.org/10.1007/s10639-022-11112-6

Iden, J., & Eikebrokk, T. R. (2013). Implementing IT service management: A systematic literature review. *International Journal of Information Management*, *33*(3), 512–523. https://doi.org/10.1016/j.ijinfomgt.2013.01.004

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press. [Database].

Jiang, M. Y.-C., Jong, M. S.-Y., Lau, W. W.-F., Chai, C.-S., & Wu, N. (2023a). Effects of automatic speech recognition technology on EFL learners' willingness to communicate and interactional features. *Educational Technology & Society*, *26*(3), 37–52. https://doi.org/10.30191/ETS.202307_26(3).0004

Jiang, M. Y.-C., Jong, M. S.-Y., Lau, W. W.-F., Chai, C.-S., & Wu, N. (2023b). Exploring the effects of automatic speech recognition technology on oral accuracy and fluency in a flipped classroom. *Journal of Computer Assisted Learning*, *39*(1), 125–140. https://doi.org/10.1111/jcal.12732

Jiang, M. Y.-C., Jong, M. S.-Y., Wu, N., Shen, B., Chai, C.-S., Lau, W. W.-F., & Huang, B. (2022). Integrating automatic speech recognition technology into vocabulary learning in a flipped English class for Chinese college students. *Frontiers in Psychology*, *13*, 902429. https://doi.org/10.3389/fpsyg.2022.902429

Kernagaran, V., & Abdullah, A. (2022). A systematic literature review of flipped learning in English as second language (ESL) context. *International Journal of Instruction*, *15*(2), 793–814. https://doi.org/10.29333/iji.2022.15243a

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, *51*(1), 7–15. https://doi.org/10.1016/j.infsof.2008.09.009

Kochem, T., Beck, J., & Goodale, E. (2022). The use of ASR-equipped software in the teaching of suprasegmental features of pronunciation: A critical review. *CALICO Journal*, *39*(3), 306–325. https://doi.org/10.1558/cj.19033

Lee, B., Plonsky, L., & Saito, K. (2020). The effects of perception- vs. production-based pronunciation instruction. *System*, *88*, 102185. https://doi.org/10.1016/j.system.2019.102185

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

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

Liu, J., Liu, X., & Yang, C. (2022). A study of college students' perceptions of utilizing automatic speech recognition technology to assist English oral proficiency. *Frontiers in Psychology*, *13*, 1049139. https://doi.org/10.3389/fpsyg.2022.1049139

Marco-Lajara, B., Martínez-Falcó, J., Millan-Tudela, L. A., & Sánchez-García, E. (2023). Analysis of the structure of scientific knowledge on wine tourism: A bibliometric analysis. *Heliyon*, *9*(2), e13363. https://doi.org/10.1016/j.heliyon.2023.e13363

McCrocklin, S. (2019a). ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation*, *5*(1), 98–118. https://doi.org/10.1075/jslp.16034.mcc

McCrocklin, S. (2019b). Learners' feedback regarding ASR-based dictation practice for pronunciation learning. *CALICO Journal*, *36*(2), 119–137. https://doi.org/10.1558/cj.34738

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System*, *57*, 25–42. https://doi.org/10.1016/j.system.2015.12.013

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery (London, England)*, *8*(5), 336–341. https://doi.org/10.1016/j.ijsu.2010.02.007

Newton, J. (2017). Pronunciation and speaking. In O. Kang, R. I. Thomson, & J. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 336–351). Routledge.

Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, *36*(1), 4–21. https://doi.org/10.1017/S0958344023000113

Ni, A., Cheung, A. C. K., & Shi, J. (2023). The impact of flipped classroom teaching on college English language learning: A meta-analysis. *International Journal of Educational Research*, *121*, 102230. https://doi.org/10.1016/j.ijer.2023.102230

Nickolai, D., Schaefer, E., & Figueroa, P. (2024). Aggregating the evidence of automatic speech recognition research claims in CALL. *System*, *121*, 103250. https://doi.org/10.1016/j.system.2024.103250

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. (2022). Directions for the future of technology in pronunciation research and teaching. *Journal of Second Language Pronunciation*, *4*(2), 182–207. https://doi.org/10.1075/jslp.17001.obr

Pennington, M. C., & Rogerson-Revell, P. (2019). *English pronunciation teaching and research*. Springer.

Rehman, I. (2021). *Real-time formant extraction for second language vowel production training (Publication Number 28417633)* [Doctoral dissertation]. Iowa State University. ProQuest Dissertations and Theses Global.

Saito, K., & Lyster, R. (2012). Effects of form-focused Instruction and corrective feedback on L2 pronunciation development of/ɹ/by Japanese learners of English. *Language Learning*, *62*(2), 595–633. https://doi.org/10.1111/j.1467-9922.2011.00639.x

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 652–708. https://doi.org/10.1111/lang.12345

Sardegna, V. G. (2022). Evidence in favor of a strategy-based model for English pronunciation instruction. *Language Teaching*, *55*(3), 363–378. https://doi.org/10.1017/S0261444821000380

Sardegna, V. G., Lee, J., & Kusey, C. (2018). Self-efficacy, attitudes, and choice of strategies for English pronunciation learning. *Language Learning*, *68*(1), 83–114. https://doi.org/10.1111/lang.12263

Shadiev, R., & Liu, J. (2023). Review of research on applications of speech recognition technology to assist language learning. *ReCALL*, *35*(1), 74–88. https://doi.org/10.1017/S095834402200012X

Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., & Akinkuolie, B. (2023). Gamification in mobile-assisted language learning: A systematic review of Duolingo literature from public release of 2012 to early 2020. *Computer Assisted Language Learning*, *36*(3), 517–554. https://doi.org/10.1080/09588221.2021.1933540

Suzuki, Y. (2023). *Practice and automatization in second language research: Perspectives from skill acquisition theory and cognitive psychology*. Taylor & Francis.

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

Vitta, J. P., Nicklin, C., & McLean, S. (2022). Effect size–driven sample-size planning, randomization, and multisite use in L2 instructed vocabulary acquisition experimental samples. *Studies in Second Language Acquisition*, *44*(5), 1424–1448. https://doi.org/10.1017/S0272263121000541

Zhu, J., Zhang, X., & Li, J. (2024). Using AR filters in L2 pronunciation training: Practice, perfection, and willingness to share. *Computer Assisted Language Learning*, *37*(5–6), 1364–1396. https://doi.org/10.1080/09588221.2022.2080716