Language Teaching (2026), **59**, 293–309
doi:10.1017/S0261444825100852

CAMBRIDGE UNIVERSITY PRESS

RESEARCH TIMELINE

Automatic speech recognition and pronunciation learning

Shannon McCrocklin[1] and John Levis[2]

[1]Linguistics, Southern Illinois University, Carbondale, IL, USA and [2]English, Iowa State University, Ames, IA, USA
Corresponding author: John Levis; Email: jlevis@iastate.edu

(Received 11 January 2025; accepted 22 June 2025)

1. Introduction

Although pronunciation is a central factor in speech intelligibility (Jenkins, 2000; Munro & Derwing, 1995, 2020), it remains a relatively neglected area of language learning because of time constraints in the classroom (Murphy & Baker, 2015), lack of teacher expertise (Huensch, 2019), and the difficulty of providing targeted feedback to diverse students (Khaustova et al., 2023). The possible application of Automatic Speech Recognition (ASR) technologies, which has the potential to provide individualized learner feedback on pronunciation, was imagined long before the technical development of ASR systems made it practicable (Wohlert, 1984*[1]). With improvements in accuracy and speaker independence of systems, interest in using ASR for pronunciation learning and teaching has grown.

This timeline traces research into the application of ASR to pronunciation teaching and learning. The process began with an initial identification of 126 articles using library searches of EBSCO Host, Linguistics and Language Behavior Abstracts, as well as follow-up searches with Google Scholar. Keywords for searches included automatic speech recognition, speech-to-text, computer-assisted pronunciation training, as well as searches seeking a combination of terms related to pronunciation as well as technology and/or computer assisted language learning. The selection process for included articles considered quality and impact of studies as well as connections between research studies across time. As a result of this iterative process, we included 46 influential articles in this timeline, which was later expanded to 50 based on reviewer feedback. Notably, many of the included articles were published in the past decade.

Automatic speech recognition works by applying computational processes to decode and transcribe oral speech input. This is done through recording oral speech input and providing an analysis of the input using a probabilistic algorithm which results in an output, usually in the form of a written text (Levis & Suvorov, 2020). Today, variants of ASR programs are found in a wide variety of applications, including personal assistants, mobile devices, video chat programs, closed captioning on videos, dictation programs for hands-free typing, and language-learning programs. As a consequence, ASR systems now play an increasingly expansive role in applied linguistics and language learning by providing feedback to learners on their spoken language and pronunciation.

With origins dating to the 1950s, ASR is a relatively young technology (Davis et al., 1952). Its modern development began in earnest in the late 1970s, and the first business-oriented ASR applications appeared in the 1980s (Rabiner & Juang, 2008). One of the main challenges facing developers of ASR was the variability of speech. Early programs were speaker-dependent; they were trained with speech samples to recognize the speech of a single speaker or a single speech variety. Even with extensive training, early programs had high error rates due to individual differences, speech rate, and

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

variations in pronunciation when words were produced in connected speech. Given these challenges, early research into ASR for L2 speech focused heavily on the accuracy of recognition (Bernstein et al., 1990*). One of the first studies to try to harness a commercial dictation program for pronunciation, Derwing et al. (2000*) examined the accuracy of transcription of a promising commercial ASR program, *Dragon Naturally Speaking*, for both L1 and L2 English speech. The study found that while the program was able to recognize L1 speech with accuracy rates of around 90%, it had much lower recognition rates for second-language (L2) speakers (∼70%) despite the fact that the L2 speakers were understood by human listeners at a much higher rate (∼95%). This study was highly influential and halted work with commercial dictation programs until later studies could document improvement in the recognition of L2 speech. Now, although some concerns remain regarding accuracy (see Inceoglu et al., 2023*), studies have found that the improved ASR recognition of L2 speech for Google Voice Typing is over 90% and is similar to that of human listeners (Johnson et al., 2024*; McCrocklin & Edalatishams, 2020*). Automatic speech recognition thus provides a useful estimate of intelligibility. Notably, improved recognition accuracy, which correlates with human listener comprehension, supports the goal of making intelligibility (in this case, the successful recognition of spoken words) rather than accentedness the emphasis in pronunciation learning (Levis, 2020).

Given concerns about the accuracy of recognition, early work in ASR for second-language learning concentrated on building bespoke Computer-Assisted Pronunciation Training (CAPT) applications that integrated ASR to enable pronunciation feedback. These CAPT programs often controlled learner utterances to enable more accurate feedback on pronunciation errors (Hincks, 2015). Some programs also incorporated other data, such as information about the learners' first languages (Moustroufas & Digilakis, 2007*), to further improve accuracy. Early examples of ASR-CAPT studies are found in Neri et al. (2002*, 2008*), which demonstrated the potential of ASR-CAPT for pronunciation learning. Interest in ASR-CAPT has continued, with examples including Bashori et al. (2022*, 2024*), which investigated both pronunciation improvement and vocabulary learning using ASR-based language-learning programs.

In the last ten years, ASR has dramatically increased in accuracy thanks to improvements based on deep learning (Cho, 2022) and the use of extensive spoken corpora (Lamel & Gauvain, 2022). These improvements reignited interest in harnessing other types of programs, such as dictation programs, as part of pronunciation learning and teaching (e.g., Liakin et al., 2015*; McCrocklin, 2016*; Mroz, 2018*), and studies highlighted the potential of dictation programs to facilitate pronunciation learning, spurring additional research that continues today (e.g., Inceoglu et al., 2024*; Johnson et al., 2024*).

Beyond accuracy and learning potential, research has also explored the experience that learners have with ASR. Although some studies shared learners' concerns about the accuracy of ASR feedback or their frustrations with low rates of recognition (McCrocklin, 2019), in general, learners like using ASR to practice their pronunciation because they get feedback about their pronunciation to help them improve (Ahn & Lee, 2016*; Neri et al., 2006*; Mroz, 2018*; Wang & Young, 2015*). The use of ASR also encourages autonomous learning (Inceoglu, 2023*; McCrocklin, 2016*). Studies using screen and video recordings of learner practice have further shown that most learners are willing to try repairing utterances when ASR indicates a possible pronunciation error (Inceoglu et al., 2024*) and can improve their ASR transcription accuracy in subsequent attempts (McCrocklin, 2019b*). Studies applying the Technology Acceptance Model (TAM) have further shown that learners and teachers are likely to continue using ASR after the conclusion of the study (Dillon & Wells, 2021*; Hsu, 2024*), indicating that they find it useful enough to make it part of their continued language learning. As studies have continued to show that ASR (both through CAPT programming or through the incorporation of generalized tools, such as dictation programs) can support pronunciation learning and improvement, researchers' questions have increasingly turned to factors that affect the success of implementation, such as the nature of the feedback provided, work arrangements, segmental targets, and length of training.

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

In general, ASR systems can provide three types of feedback. First, they can provide an overall score for learners' pronunciation, for example, a Goodness of Pronunciation (GOP) score (see Witt and Young, 1997*). However, GOP scores are opaque, requiring learners to determine where their pronunciation does not succeed in order to increase their scores. Automatic speech recognition feedback can also be provided in the form of written text, in which errors in transcription may suggest pronunciation errors. However, this type of feedback is also indirect, and mistranscriptions may not help learners understand the patterns of errors in their speech nor how to improve. Finally, ASR can be used in conjunction with explicit feedback on particular phonemic errors to highlight mispronunciations and provide explanatory feedback. The meta-analysis by Ngo et al. (2024*) showed that while implicit feedback was useful and could support learning gains, explicit feedback was more beneficial for learning. Still, more work is needed to explore the effectiveness of various ASR-feedback options.

Another area of investigation has been work arrangements for ASR practice, such as whether the work is done individually, in pairs or groups, or in conjunction with teacher-training. Although most early research assumed learners would work individually with ASR programs, and Elimat & AbuSeileek (2014*) found an advantage to students working individually, studies have increasingly found that learner support (practice with a peer) may be able to increase the efficacy of such practice (Dai & Wu, 2023*; Evers & Chen, 2021*). Learners also expressed greater satisfaction working with ASR when they had teacher support (Liu et al., 2022*). It is likely that there is a link between the feedback quality of the ASR program and the need for additional support, either through a peer or teacher. Learners may derive greater benefits from support when using programs that provide indirect or implicit feedback, as learners may struggle to interpret implicit feedback on their own. This should be a target for further research.

Several studies have shown that ASR training has uneven effects across segmental pronunciation targets. McCrocklin (2019a*) examined learner improvement across a range of English vowels and consonants, finding that some segmentals (for example /ɪ/ and /i/) showed little to no improvement. Inceoglu et al. (2020*) and Guskaroska (2020) also examined learner improvement for vowel sounds in English, finding that some vowel pronunciations improved while others did not. Future research may explore whether there are consistent differences, which would suggest that some segmentals are more suitably trained through ASR practice than others.

Due to practical constraints, many ASR studies have relied on short training windows. A recent meta-analysis (Ngo et al., 2024*), however, suggests that longer and more consistent use of ASR leads to greater improvement. Their analysis showed that studies involving four weeks of training or less were rarely effective, while studies employing medium (5–8 weeks) or long training durations (9 weeks or longer) led to substantial improvement. Studies that seek longer training times, such as Hsu (2024*) and Nickolai (2024*), will help the field better understand the ultimate impacts of sustained ASR training and the extent of improvement possible with ASR practice.

Looking to the future, ASR programs will continue to grow in influence within language learning in general and for pronunciation learning in particular. Automatic speech recognition continues to improve, not only in general, but specifically in recognizing and assessing L2 speech (Çalık et al., 2024; Yan et al., 2024*). Ultimately, the gold standard of ASR for pronunciation improvement (Derwing & Munro, 2015) occurs when listeners understand the spontaneous speech of L2 speakers more successfully. Programming that integrates artificial intelligence (AI) will increasingly allow learners to engage in spontaneous speech practice while getting pronunciation feedback on features affecting intelligibility; generative AI will likely blur the lines between CAPT, which can offer explicit feedback, but often offers little to no flexibility for learners to actively participate in creating and producing their own sentences, and repurposed commercial programs (e.g., dictation programs), which can only provide implicit feedback through a written transcript, but allow learners flexibility and control to create and practice their own learning materials. Currently, programs like *ELSA Speak* and *Gliglish* are debuting options for flexible, communicative practice through role-play conversations

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

that make explicit pronunciation feedback available. Newer programs, however, may have language-learning goals that focus on other language learning skills (see Bashori et al., 2024*) but do not support sustained focus on pronunciation. Ideally, future research will continue to closely examine the pronunciation learning enabled through these technologies and continue to find ways to better support learners' pronunciation improvement.

Our exploration of the history of research in ASR for pronunciation learning and teaching identified five major themes. Coding advancements (A) have largely taken place apart from pronunciation learning and teaching but have allowed for the current usefulness of ASR for pronunciation learning and teaching. Evaluations of programs in terms of accuracy and affordances (B) for learning continue to play a role in research, but these areas are less important now than in the past as the programs have been established as sufficiently accurate to be useful for pronunciation learning. In regard to types of programs (C), early emphasis on specially created programs has been replaced by emphasis on repurposing programs built for other purposes (e.g., dictation programs). Future developments of programs integrating AI or focusing on suprasegmental features may again change the direction of research in this area. Research has focused heavily on learner improvement (D), or how and where learners improve. While most studies have focused on improvements of segmental pronunciation, some studies have focused on larger issues of intelligibility or comprehensibility. Research on ASR for suprasegmental training has been rare but may become more common if or when commercial programs become available that can address suprasegmentals (see Kochem et al., 2022*). Research has also explored the learner experience (E), including how learners use the tools as well as the degree to which learners find ASR enjoyable and valuable as a language learning tool.

A. Coding advancements
B. Program evaluations

   B1. Accuracy of ASR
   B2. Affordances for learning

C. Type of program

   C1. Assessment
   C2. CAPT programs
   C3. Dictation programs

D. Learner improvement

   D1. Segmental accuracy changes
   D2. Intelligibility/comprehensibility

E. Learning experience

Note

1. Asterisk indicates full reference can be found in the timeline itself.

References

Çalık, Ş. S., Küçükmanisa, A., & Kilimci, Z. H. (2024). A novel framework for mispronunciation detection of Arabic phonemes using audio-oriented transformer models. *Applied Acoustics*, *215*, 109711. https://doi.org/10.1016/j.apacoust.2023.109711

Cho, K. (2022). Deep learning. In R. Mitkov (Ed.), *The Oxford handbook of computational linguistics* (pp. 359–414). Oxford University Press.

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

Davis, K. H., Biddulph, R., & Balashek, S. (1952). Automatic recognition of spoken digits. *The Journal of the Acoustical Society of America*, *24*(6), 637–642. https://doi.org/10.1121/1.1906946

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins

Guskaroska, A. (2020). ASR-dictation on smartphones for vowel pronunciation practice. *Journal of Contemporary Philology*, *3*(2), 45–61. https://doi.org/10.37834/jcp2020045g

Hincks, R. (2015). Technology and learning pronunciation. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 505–519). John Wiley & Sons.

Huensch, A. (2019). Pronunciation in foreign language classrooms: Instructors' training, classroom practices, and beliefs. *Language Teaching Research*, *23*(6), 745–764. https://doi.org/10.1177/1362168818767l8

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

Khaustova, V., Pyshkin, E., Khaustov, V., Blake, J., & Bogach, N. (2023, November). CAPTuring accents: An approach to personalize pronunciation training for learners with different L1 backgrounds. In *International conference on speech and computer* (pp. 59–70). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-48312-7_5

Lamel, L., & Gauvain, J.-L. (2022). Speech recognition. In R. Mitkov (Ed.), *The Oxford handbook of computational linguistics* (pp. 770–788). Oxford University Press.

Levis, J. (2020). Revisiting the intelligibility and nativeness principles. *Journal of Second Language Pronunciation*, *6*(3), 310–328. https://doi.org/10.1075/jslp.20050.lev

Levis, J., & Suvorov, R. (2020). Automatic speech recognition. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Wiley. https://doi.org/10.1002/9781405198431.wbeal0066.pub2

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

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (2020). Foreign accent, comprehensibility and intelligibility, redux. *Journal of Second Language Pronunciation*, *6*(3), 283–309. https://doi.org/10.1075/jslp.20038.mun

Murphy, J. M., & Baker, A. A. (2015). History of ESL pronunciation teaching. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 36–65). Wiley-Blackwell.

Rabiner, L., & Juang, B. H. (2008). Historical perspective of the field of ASR/ NLU. In J. Benesty, M. M. Sondhi, & Y. A. Huang (Eds.), *Springer handbook of speech processing* (pp. 521–538). Springer.

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

Year	References	Annotations	Theme
1984	Wohlert, H. S. (1984). Voice input/output speech technologies for German language learning. *Die Unterrichtspraxis/Teaching German, 17*(1), 76–84. https://doi.org/10.2307/3530426	An early example, **Wohlert** used newly affordable technology, Vet-2, to provide German practice for L2 learners in a language laboratory. Vet-2 used templates based on the instructor's model to match learners' practice productions to text. Although the study included notable weaknesses, participants showed more enthusiasm to study with Vet-2 than at home, despite some technical difficulties, and had higher post-test scores, highlighting the potential of speech technologies as part of CAPT.	D, E
1990	Bernstein, J., Cohen, M., Murveit, H., Rtischev, D., & Weintraub, M. (1990). Automatic evaluation and training in English pronunciation. In *International congress on spoken language processing* (Vol. 90, pp. 1185–1188). https://doi.org/10.21437/icslp.1990-313	A pioneering attempt to build a program using ASR (*Decipher*) to score L2 pronunciation automatically. To evaluate the program, **Bernstein and colleagues** collected audio recordings from Japanese L1 speakers reading English sentences. The program used an HMM ASR to generate scores which were compared to expert listeners' segmental and prosodic ratings. The correlation between the program and expert listeners was high ($>$0.8) when using the program's assessment of the full spoken sentence. The authors concluded that automated scoring of L2 pronunciation was possible, opening the door to future research.	A, B1, C1
1993	Hiller, S., Rooney, E., Laver, J., & Jack, M. (1993). SPELL: An automated system for computer-aided pronunciation teaching. *Speech Communication, 13*(3–4), 463–473. https://doi.org/10.1016/0167-6393(93)90045-m	Calling their program, *SPELL*, a second-generation technology, the authors applied ASR capabilities to pilot a CAPT program that could support learning of intonation, rhythm, and vowel quality. The study described *SPELL* and reported preliminary evaluations by instructors who identified several strengths as well as areas for future development. Instructors saw potential for the system to be useful once further developed. *SPELL* gained recognition, serving as a useful model for future ASR-CAPT development.	B2, C2
1997	Witt, S. & Young, S. (1997). Language learning based on non-native speech recognition. *Eurospeech,* 1–4. https://doi.org/10.21437/eurospeech.1997-227	Noting that previous ASR systems assessed phone quality based on a percentage of accuracy within a word (e.g., HILLER ET AL., 1993[1]), **Witt and Young** developed a system that adapted HMM to assess the quality of individual phones using Goodness of Pronunciation (GOP) that allowed feedback on accurate and inaccurate phones in the learners' inventory. Their model was effective in detecting mispronunciations, a first step in being able to provide specific feedback and individualized training for learners as part of CAPT.	A

(*Continued*)

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

(*Continued.*)

Year	References	Annotations	Theme
1999	Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. *System, 27*, 49–64. https://doi.org/10.1016/s0346-251x(98)00049-9	Similar to WOHLERT (1984), **Coniam** saw potential to harness commercial programs for language learning. He was the first to assess a dictation program, *Dragon Naturally Speaking*, for L2 speech. Coniam compared the ASR-transcript accuracy for L2 (Chinese L1) and L1 English speakers. Despite substantial training using learners' voices, the study found substantially lower transcript accuracy for L2 speech. Coniam concluded that ASR-dictation software was not ready to be used for L2 speaker training or assessment.	B1, C3
1999	Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning and Technology, 2*(2), 62–76. http://dx.doi.org/10125/25043	After acknowledging increases in accuracy and availability of speech recognition technologies, **Eskenazi** used principles established for pronunciation teaching to evaluate an ASR CAPT system, *FLUENCY,* for pronunciation learning. She noted positive aspects such as the potential to provide corrective feedback, while highlighting areas for continued development, e.g., the need to increase students' active participation in sentence production (creating and using their own sentences for practice), an issue that still persists in many CAPT systems today.	B2, C2
2000	Derwing, T., Munro, M., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly, 34*(3), 592–603. https://doi.org/10.2307/3587748	Expanding on CONIAM (1999), **Derwing et al.** also examined *Dragon Naturally Speaking*'s transcription accuracy for English, but notably compared the ASR transcript to human-listener transcripts and ratings. The study found a substantial drop in ASR accuracy for L2 speech and no statistically significant correlations with human measures. Taken together with CONIAM's (1999) findings, Derwing and colleagues concluded dictation programs were currently insufficient for use in ESL education or assessment. Their conclusions stalled research into dictation for language education for 15 years (see LIAKIN ET AL. [2015] and the study's eventual replication in MCCROCKLIN & EDALATISHAMS [2020]) refocused the field on the promise of CAPT.	B1, C3
2000	Franco, H., Neumeyer, L., Digalakis, V., & Ronen, O. (2000). Combination of machine scores for automatic grading of pronunciation quality. *Speech Communication, 30*, 121–130. https://doi.org/10.1016/s0167-6393(99)00045-x	Also using the ASR *Decipher* as a starting point, **Franco and colleagues** built on BERNSTEIN ET AL. (1990)'s work by exploring various methods of pronunciation scoring to create better predictive alignment with human expert ratings. The study argued that a combination of scores, rather than any single measure, better predicted human expert ratings and nonlinear mapping functioned better than linear mapping, advancing ASR recognition and scoring of L2 speech considerably.	A, B1

(*Continued*)

(*Continued.*)

Year	References	Annotations	Theme
2000	Kawai, G. & Hirose, K. (2000). Teaching the pronunciation of Japanese double-mora phonemes using speech recognition technology. *Speech Communication, 30*, 131–143. https://doi.org/10.1016/s0167-6393(99)00041-2	Using WITT & YOUNG's (1997) GOP algorithm, the authors built a CAPT system teaching Japanese double-mora phonemes. After initial system evaluations, students practiced with minimal pairs and received scores and feedback based on vowel duration. The authors used pre-/post- testing to examine learner improvement, both on trained and untrained stimuli. The study showed notable gains in learner accuracy that transferred to new, untrained stimuli and was one of the first studies to substantiate claims of ASR CAPT's learning potential.	B1, C2, D1
2000	Witt, S., & Young, S. (2000). Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication, 30*, 95–108. https://doi.org/10.1016/s0167-6393(99)00044-8	Expanding on WITT & YOUNG (1997) the authors further enhanced scoring of L2 pronunciation in CAPT programming. While still using a Goodness of Pronunciation approach, they were able to improve their model with a likelihood-based method that incorporated data from the speakers' L1s and phone dependent thresholds based on averages of human judgements.	A, B1
2002	Neri, A., Cucchiarini, C., Strik, H., & Boves, L.W.J. (2002). The pedagogy-technology interface in Computer Assisted Pronunciation Training. *Computer Assisted Language Learning, 15*(5), 441–467. https://doi.org/10.1076/call.15.5.441.13473	Building on ESKENAZI (1999), Neri and colleagues performed an appraisal of various CAPT programs following pedagogical principles. The study highlights that many commercial systems focused on incorporating 'technological novelties' while ignoring features more likely to support actual learning and improvement. The analysis and recommendations were influential in setting guidelines for future approaches.	B2, C2
2006	Neri, A., Cucchiarini, C., & Strik, H. (2006). ASR-based corrective feedback on pronunciation: Does it really work? In *Proceedings of the ISCA Interspeech 2006*, Pittsburgh, PA (pp. 1982–1985). https://doi.org/10.21437/interspeech.2006-543	The study compared Dutch learning outcomes and learner perspectives when learners practiced with either an ASR-CAPT program or a non-ASR CAPT program. Learners appreciated both CAPT training programs, but when expert raters analyzed pre- and post-test recordings (read speech featuring all sounds in Dutch) with a holistic rubric of overall segmental production, the ASR-CAPT training group made larger gains in their pronunciation accuracy.	C1, D2, E
2007	Moustroufas, N., & Digalakis, V. (2007). Automatic pronunciation evaluation of foreign speakers using unknown text. *Computer Speech and Language, 21*, 219–230. https://doi.org/10.1016/j.csl.2006.04.001	While previous work on ASR assessment of learner speech focused on CAPT programs in which the learner's speech was predetermined and controlled (e.g., WITT & YOUNG [2000]), this study analyzed an approach to improve automated scoring of L2 speech without knowledge of the uttered text. The authors found that combining coding based on both the target and source languages allowed the program to provide scores that were closely aligned with human raters.	A, B1, C1

(*Continued*)

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

(*Continued.*)

Year	References	Annotations	Theme
2008	Cincarek, T., Gruhn, R., Hacker, C., Nöth, E., & Nakamura, S. (2008). Automatic pronunciation scoring of words and sentences independent from the non-native's first language. *Computer Speech and Language, 23*, 65–88. https://doi.org/10.1016/j.csl.2008.03.001	Working toward different goals than MOUSTROUFAS & DIGILAKIS (2007), **Cincarek and colleagues** built on WITT & YOUNG (2000)'s developments, which modeled ASR after L1 data, by improving scoring accuracy for controlled L2 speech within CAPT programs when the first language was unknown. Using an HMM chain model, the authors were able to provide automated scoring almost equal to average human evaluations.	A, B1, C2
2008	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	Building on previous research on adult learners (e.g., NERI ET AL., 2006), this study looked at young learners. The study examined English pronunciation improvement of 11-year old Italian L1 learners, comparing face-to-face training to ASR training with the CAPT system, *PARLING*. Pre-/post-testing showed both groups improved their pronunciation substantially, but ASR users spent less time in training.	C2, D2
2014	Elimat, A. K., & AbuSeileek, A. F. (2014). Automatic speech recognition technology as an effective means for teaching pronunciation. *JALT CALL Journal, 10*(1), 21–47. https://doi.org/10.29140/jaltcall.v10n1.j166	Similar to NERI ET AL. (2008) **Elimat and AbuSeileek** looked at child learners, using *Tell Me More*, which integrated ASR. It expanded on previous works by including three experimental groups (individual work, pair-work, group work with ASR) in addition to a control group. Results showed all ASR-training groups achieved greater improvement than the control group, with a particular advantage for the group working individually.	C2, D2, E
2015	Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal, 32*(1), 1–25. https://doi.org/10.1558/cj.v32i1.25962	Ending the 15-year lull in research focused on dictation programs following DERWING ET AL. (2000), this study used *Dragon Dictation* to examine improvement of the French vowel /y/ for adult learners in three groups, a pronunciation-focused ASR group, a pronunciation-focused instructor-led group, and a control group with weekly conversation meetings. Only the ASR group significantly improved.	C3, D1
2015	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*, 493–504. https://doi.org/10.1111/jcal.12079	This was the first study to look at the effectiveness of formative and summative feedback in various modalities as part of ASR-CAPT training for both young and adult learners. Findings showed formative feedback benefitted learners' pronunciation more, while summative feedback was more likely to be used by learners to track progress. Adult learners reported benefitting from audio feedback, while younger learners appreciated more explicit textual feedback.	C2, D1, E

(*Continued*)

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

(*Continued.*)

Year	References	Annotations	Theme
2016	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	**Ahn and Lee** focused on learner experiences through surveys of Korean middle-school learners following experiences with *Speaking English 60 Junior*, a mobile speaking application integrating ASR to facilitate practice tasks. Students appreciated drills and shadowing activities with ASR for providing chances to speak, noting that drills provided repetition to support remembering and improvement over iterations. The study also highlighted the potential of ASR to support noticing of pronunciation errors during communicative practice.	E
2016	McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System, 57*, 25–42. https://doi.org/10.1016/j.system.2015.12.013	Focusing on one aspect of learner experience, autonomy, **McCrocklin** followed the autonomous learning beliefs and behaviors of three groups over three weeks. While the control group with traditional face-to-face training showed no change in beliefs, both ASR groups significantly increased beliefs of their autonomy. The hybrid group, with half of its work through ASR, reported more time spent on autonomous pronunciation learning and more use of ASR, than the group that only received strategy training in ASR.	C3, E
2018	Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals, 51*(3), 1–21. https://doi.org/10.1111/flan.12348	Building on LIAKIN ET AL. (2015) and MCCROCKLIN (2016), **Mroz** focused on autonomous French learning with dictation in *Gmail*. The study examined learner experiences through a qualitative approach utilizing interviews. Learners reported that the practice with ASR made them more aware of French spelling and its correspondence to pronunciation. They also reported appreciating that ASR gave them a sense of how human listeners might perceive their speech.	E
2019	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	Building on LIAKIN ET AL. (2015) which focused on acquisition of a single sound in French, **McCrocklin** examined learners' (L2 English) improvement for a range of consonant and vowel segments from three weeks of pronunciation training of two groups, one conventional face-to-face and one hybrid training group receiving half of their instruction with *Windows Speech Recognition*. Both groups were assessed pre- and post-workshop on production of nine segmentals and both improved in accuracy showing ASR practice may be equivalent to instructor-led practice.	C3, D1

(*Continued*)

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

(*Continued*.)

Year	References	Annotations	Theme
2019	McCrocklin, S. (2019b). Dictation programs for second language pronunciation learning: Perceptions of the transcript, strategy use and improvement. *Konin Language Studies, 7*(2), 137–157. https://doi.org/10.30438/ksj.2019.7.2.2	This study was the first to examine learners' real-time use of ASR during practice, using think-aloud protocols and tracking of practice behaviors as L2 English learners used *Google Voice Typing* for practice and feedback. Results showed learners used the transcript to identify words with probable errors and hypothesized about segmental errors or articulatory features causing errors. Learners relied on a range of strategies to improve their pronunciation and were able to improve the transcript accuracy in subsequent attempts, although the benefits of repetition decreased after three attempts.	C3, D1, E
2020	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. https://doi.org/10.18823/asiatefl.2020.17.3.5.824	Building on LIAKIN ET AL. (2015) and MCCROCKLIN (2019a), **Inceoglu et al.** used acoustic analysis to better examine segmental improvement of two vowel minimal pairs. It followed an experimental group (EG, no control group) of Korean EFL learners' improvement of four vowels using ASR. EG read 28 minimal pair sentences at pre-test and post-test, with six out-of-class practice sessions. Although participants liked using ASR as a tool for pronunciation practice, acoustic analysis showed improvement on only one of the four vowels analyzed.	C3, D1, E
2020	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	In this study, second-semester learners of Spanish used ASR (*iSpraak*) or received instructor-led pronunciation practice over a 15-week term. On the basis of pre-test and post-test reading, ASR was better suited for targeted segmental learning, but instructor-led learning was better for comprehensibility improvement, suggesting that the two approaches were complementary.	B1, D
2020	McCrocklin, S., & Edalatishams, I. (2020). Revisiting popular speech recognition software for ESL speech. *TESOL Quarterly, 54*(4), 1086–1097. https://doi.org/10.1002/tesq.3006	A partial replication of DERWING ET AL. (2000), this study examined the accuracy of ASR (Google Voice Typing, GVT) for L2-accented speech. GVT recognized accented speech at over 90% and was similar to the recognition rate levels of human listeners indicating that at least some ASR programs had become equivalent to human listeners at recognizing L2 speech.	B1, C3
2020	Mroz, A. (2020). Aiming for advanced intelligibility and proficiency using mobile ASR. *Journal of Second Language Pronunciation, 6*(1), 12–38. https://doi.org/10.1075/jslp.18030.mro	Building on MROZ (2018), **Mroz** compared two university courses on L2 French: one was focused on proficiency goals and the other on pronunciation using ASR. Improvement was measured for French intelligibility and proficiency with read-aloud and semi-spontaneous speech. Participants exposed to ASR, regardless of the course, significantly improved in their intelligibility, but not in proficiency. The impact of ASR was strengthened when learners were taught spelling patterns as part of the pronunciation-focused course.	C3, D2

(*Continued*)

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

(*Continued.*)

Year	References	Annotations	Theme
2020	Tejedor-García, C., Escudero-Mancebo, D., Cámara-Arenas, E., González-Ferreras, C., & Cardeñoso-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	Focusing on student learning outcomes, **Tejedor-García et al.** used and evaluated an adaptive CAPT tool integrating ASR to facilitate feedback on minimal pair production. Using pre- and post-tests, this well-designed study examined impacts of a 3-week training program with either CAPT or classroom practice, showing significantly higher scores for the CAPT training group on the post-test. The study also highlighted alignment between human and ASR assessments of the post-test.	B1, C2, D1
2021	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	Using the Technology Acceptance Model (TAM), **Dillon and Wells** examined experiences and beliefs of 211 Korean learners of English who practiced with ASR over 14 weeks of self-study. Although the majority of participants found the tool useful for practice, the study noted learner support for the use of ASR for testing purposes was mixed.	C1, E
2021	Evers, K. & Chen, S. (2021). Effects of Automatic Speech Recognition software on pronunciation for adults with different learning styles. *Journal of Educational Computing Research, 59*(4), 669–685. https://doi.org/10.1177/0735633120972011	Expanding on MCCROCKLIN (2016, 2019a) and MROZ (2018), this study was the first to focus on individual differences in working with ASR. Using three groups (control, ASR with peer correction, and ASR alone), the study examined pronunciation improvement using pre- and post-tests, highlighting the usefulness of ASR when used in conjunction with peer correction. Considering individual differences through a paper-based learning styles inventory, the study used a two-way ANCOVA to show learning preferences had a notable impact on learning outcomes. Interestingly, learners with verbal styles made less progress with ASR practice alone, benefitting more from the integration of peer correction.	B2, D
2021	Oh, E. Y., & Song, D. (2021). Developmental research on an interactive application for language speaking practice using speech recognition technology. *Educational Technology Research and Development, 69*(2), 861–884. https://doi.org/10.1007/s11423-020-09910-1	Echoing the work of ESKENAZI (1999), **Oh and Song** sets forth a principled approach to building new CAPT tools that integrate ASR using language learning theory to inform initial development, developing a functional prototype, and conducting a cycle of testing and revising to reach a workable version.	B2, C2
2021	Spring, R., & Tabuchi, R. (2021). Assessing the practicality of using an automatic speech recognition tool to teach English pronunciation online. *Journal of English Teaching through Movies and Media*, *22*(2), 93–104. https://doi.org/10.16875/stem.2021.22.2.93	Using an ASR-CAPT tool, *NatTos*, as part of a fully online course, **Spring and Tabuchi** examined effects of training on learner intelligibility and fluency. Using a read-aloud task, the pre- and post-test scores showed improvements to intelligibility following the ASR practice for mid to lower proficiency learners. However, there were small declines in most learners' fluency suggesting learners' attention to accuracy detracted from attention to fluency.	C2, D2

(*Continued*)

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

(*Continued.*)

Year	References	Annotations	Theme
2022	Bashori, M., van Hout, 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, 37*(5–6), 1335–1363. https://doi.org/10.1080/09588221.2022.2080230	This study explored impacts of ASR training on pronunciation and vocabulary improvement for two groups of Indonesian students. They received training with different websites that incorporated ASR. Using a phonetic analysis tool, along with human expert analysis, pronunciation improvement was measured for 40 words. Compared to the control group, learners in the two experimental groups with ASR showed greater improvement in both pronunciation accuracy and vocabulary knowledge.	C2, D1
2022	Kochem, T., Beck, J., & Goodale, E. (2022). The use of ASR-equipped software in the teaching of suprasegmental features of pronunciation. *CALICO Journal, 39*(3), 306–325. https://doi.org/10.1558/cj.19033	**Kochem et al.** examines research trends and programs that integrate speech recognition to train learners in suprasegmentals. While noting ASR has been able to support assessment goals for suprasegmental features (e.g., fluency and word boundaries), most programs are not commercially available, limiting their use and research for teaching and learning purposes. The researchers call for development of commercial programs to support the use of ASR for teaching and learning suprasegmentals.	B2, C1, C2
2022	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	Similar to DILLON & WELLS (2021), this study used TAM to compare learner experiences. While both groups received training with an ASR program, the EG which also received teacher help showed greater satisfaction with the ASR tool.	C2, E
2022	Van Lieshout, C., & Cardoso, W. (2022). Google Translate as a tool for self-directed language learning. *Language Learning and Technology, 26*(1), 1–19. http://hdl.handle.net/10125/73460	This study explored the use of *Google Translate* for learning Dutch phrases and their pronunciation. The study included a pre-test, post-test, and delayed post-test. The learners were able to learn the phrases on a short-term basis and were able to combine the use of ASR and TTS (Text-to-Speech) in varied ways to help them learn.	C3, D, E
2023	Cámara-Arenas, E., Tejedor-Garcia, C., Tomas-Vásquez, C. J., & Escuder-Mancebo, D. (2023). Automatic pronunciation assessment vs. automatic speech recognition: A study of conflicting conditions for L2-English. *Language Learning and Technology, 27*(1), 1–19. https://hdl.handle.net/10125/73512	**Cámara-Arenas et al.** examined the use of Google ASR for automatic pronunciation assessment. Using recognition rates of both L1 and foreign-accented speech for sets of minimal pairs, including rare ones, the authors' findings raised concerns that Google ASR cannot meet two conditions (intrinsic and derived) that would allow it to function as a valid assessment.	B1, C1
2023	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	Building on EVERS & CHEN (2021), **Dai and Wu** looked at pronunciation learning for three groups receiving either (1) peer feedback only, (2) ASR feedback only, or (3) both peer and ASR feedback. All three groups improved their pronunciation, but the two groups receiving peer feedback outperformed the ASR only group.	C3, D1, D2

(*Continued*)

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

(*Continued.*)

Year	References	Annotations	Theme
2023	Inceoglu, S. (2023). Developing pronunciation learner autonomy with Automatic Speech Recognition and shadowing. In S. McCrocklin (Ed.), *Technological resources for second language pronunciation learning and teaching: Research-based approaches* (pp. 171–192). Lexington Books.	Similar to McCrocklin (2016), this study broadly focused on learner autonomy to understand learner experiences in an online course using shadowing and ASR for individual practice. In addition to finding that learners found ASR useful in monitoring their speech, the study echoed findings of Spring & Tabuchi (2021), as shadowing led to much faster speech delivery than ASR practice. This continues to raise questions about ASR's impact on fluency.	C3, D, E
2023	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	This study looked at intelligibility measures for L2 speech as evaluated by ASR and L1 speakers focusing on the pronunciation of /i–ɪ/ and /æ–ɛ/ in words and in sentences. ASR and humans were both more successful in transcribing sentences than in transcribing words. Despite similarities in error types, however, the study highlighted that ASR recognition mirrored human listening in only one of four Taiwanese speakers, raising questions about the effects of individual L2 speech differences on recognition.	B1, C3
2023	Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., & Wu, N. (2023). 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	In line with Inceoglu (2023), **Jiang et al.** looked at the impacts of ASR practice on both oral accuracy and fluency for 160 students using a flipped classroom approach in their first year of study. Improvement was measured at four different times during the term for an experimental group and control group finding that phonological accuracy improves with the use of ASR, but accuracy improvements may have negatively affected fluency gains.	B1, C3, D
2023	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	Continuing to build on our understanding of the importance of peer feedback with ASR (e.g., Evers & Chen [2021] and Dai & Wu [2023]), **Sun** suggests that a combination of ASR plus peer feedback may support greater improvement than teacher-led instruction.	C3, D2
2024	Ngo, T. H., Chen, H. J., Lai, K. W. (2024). The effectiveness of automatic speech recognition for ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 1–18. https://doi.org/10.1017/s0958344023000113	This meta-analysis of 15 ASR studies found that ASR as part of instruction improved segmentals more than suprasegmentals, that longer treatments were better, that ASR is more effective with peer practice, and that explicit feedback was better than implicit (but both supported learner improvement).	D
2024	Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024). 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	Building on Bashori et al. (2022), **Bashori et al**. examined the effects of two specially developed ASR-equipped websites for two different groups of Indonesian English learners, one providing global pronunciation feedback and the other specific feedback on particular errors. Both groups improved their pronunciation, but the ASR program providing specific feedback led to greater improvement.	D1, D2

(*Continued*)

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

(*Continued.*)

Year	References	Annotations	Theme
2024	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	Similar to Dillon & Wells (2021) and Liu et al. (2022), **Hsu** reports on students' perceptions of the usefulness of ASR after two semesters of study in which ASR was used within a CAPT framework. The results were interpreted through the TAM and suggested that the students found ASR-based practice useful for working on their pronunciation.	B2, D1, E
2024	Inceoglu, S., Chen, W. H., & Lim, H. (2024). Monitoring student behavior in autonomous automatic speech recognition-based pronunciation practice. *System, 124*, 103387. https://doi.org/10.1016/j.system.2024.103387	Similar to McCrocklin (2019b), this study looked at how learners interact with ASR systems during practice. Using screen recordings to look at how learners respond to mistranscribed words, **Inceoglu and colleagues** examined correlations between behaviors and learners' attitudes toward pronunciation and motivation at the beginning of the study. Attitudes correlated with a willingness to repair mispronunciations, but motivation was not correlated to learner behavior.	E
2024	Johnson, C., Cardoso, W., Zuercher, B., Brannen, K., & Springer, S. (2024). Assessing pronunciation using dictation tools: The use of Google Voice Typing to score a pronunciation placement test. *Journal of Second Language Pronunciation, 10*(1), 10–34. https://doi.org/10.1075/jslp.23033.joh	Building on McCrocklin & Edalatishams (2020), **Johnson et al.** evaluated the accuracy of Google Voice Typing for L2 speech, noting strong correlations between in-house, spoken-language test results, and scores based on GVT transcriptions. In a finding contrary to Camara et al. (2023), Johnson and colleagues suggest that the use of ASR can help provide information for an accessible and affordable automated assessment of spoken language ability.	B1, C1, C3
2024	Nickolai, D. (2024). Quantifying the impact of ASR-based instruction: What does the iSpraak platform learner data show? *The EuroCALL Review, 31*(1), 16–23. https://doi.org/10.4995/eurocall.2024.20221	One of the first studies to provide a large data approach, **Nickolai** reports learner improvement across eight years of data (using 18,061 records) from iSpraak (also used in Garcia et al. [2020]), an ASR tool for multiple languages. Quantitative results indicate gains in intelligibility for users engaged in multiple rounds of ASR practice. Findings also showed the degree of improvement may not be the same across various L2s.	C2, D2

(*Continued*)

(*Continued.*)

Year	References	Annotations	Theme
2024	Nickolai, D., Schaefer, E., & Figueroa, P. (2024). Aggregating the evidence of automatic speech recognition claims in CALL. *System, 121*, 1–41. https://doi.org/10.1016/j.system.2024.103250	Taking a complementary approach to Ngo et al. (2024), this review article sought to establish the veracity of five claims about ASR made in research through analysis of 50 published research studies. Their analysis found that all claims had support, noting compelling evidence that ASR can signal mispronunciations to learners, produce measurable gains in learner speech, and create positive learning experiences. Evidence that ASR evaluates speech like humans and that ASR is at least as effective as teacher-led instruction was compelling.	B1, C, D, E
2024	Yan, B. C., Li, J. T., Wang, Y. C., Wang, H. W., Lo, T. H., Hsu, Y. C., Chao, W. C., & Chen, B. (2024, August). An effective pronunciation assessment approach leveraging hierarchical transformers and pre-training strategies. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, Volume 1: Long Papers* (pp. 1737–1747). https://doi.org/10.18653/v1/2024.acl-long.95	This study examined coding advancements featuring innovative approaches to pre-training for modeling and hierarchical methods for automatic pronunciation assessment. The approach allowed the program to assess multiple pronunciation features at once while maintaining program functionality and accuracy.	A, B1, C1

[1]Small capitalised work is mentioned elsewhere in the timeline.

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press

Shannon McCrocklin is an Associate Professor in the School of Languages and Linguistics at Southern Illinois University. She earned her M.A. in TESOL from the University of Illinois and her Ph.D. in Applied Linguistics and Technology from Iowa State University. Her research focuses on the intersection of CALL and second language pronunciation learning and teaching, with a primary emphasis on automatic speech recognition. She is the editor of *Technological resources for second language pronunciation learning and teaching: Research-based approaches* (2022, Bloomsbury Publishing).

John Levis is Distinguished Professor of Applied Linguistics and TESL at Iowa State University. He is the author of *Intelligibility, oral communication, and the teaching of pronunciation* (2018, Cambridge University Press) as well as the co-author of *Second language pronunciation: Bridging the gap between research and practice* (2022, Wiley Blackwell). He co-edited the *Handbook of English pronunciation* (2015, Wiley Blackwell) as well as *Social dynamics in second language accent* (2014, DeGruyter Mouton). He is the founder of an annual conference, Pronunciation in Second Language Learning and Teaching, held since 2009, and the founding editor of the *Journal of Second Language Pronunciation*.

Cite this article: McCrocklin, S., & Levis, J. (2026). Automatic speech recognition and pronunciation learning. *Language Teaching*, 59(3), 293–309. https://doi.org/10.1017/S0261444825100852

https://doi.org/10.1017/S0261444825100852 Published online by Cambridge University Press