TEACHING AND LEARNING PRONUNCIATION WITH TECHNOLOGY: CURRENT POSSIBILITIES, PEDAGOGICAL RECOMMENDATIONS, AND DIRECTIONS FOR THE FUTURE[1]

Jonás Fouz-González, University of Murcia

Technology has always played a central role in pronunciation teaching and learning. Computer Assisted Pronunciation Training (CAPT) research provides valuable insights into the potential of various tools and techniques. However, publications sometimes overlook how these fit into the broader framework of pronunciation instruction. Hence, while studies offer important information regarding what specific technologies can do, their contributions to the teaching-learning process beyond those specific studies are not always clear. Moreover, the potential of technology is sometimes discussed in rather general terms, without specifying the tools or pedagogical implementations that can help teachers and learners achieve concrete aims. Since the effectiveness of technology depends on the pedagogical approach adopted, it is crucial to consider not only each tool's affordances, but also learners' needs at different stages of the teaching-learning process and the ways in which technology can support them. This article reviews how technology can enhance pronunciation practice, drawing on empirical findings from the field of CAPT. It then provides recommendations for effectively integrating technology into pronunciation instruction. Finally, it discusses the future of CAPT, challenges that need to be addressed, and recommendations for advancing future research and strengthening the connection between research and practice.

INTRODUCTION

Researchers have been exploring the potential of technology to facilitate the teaching and learning of pronunciation for decades. However, despite the extensive body of research on the topic, it is difficult to be fully aware of all the possibilities technology offers for pronunciation practice and of how the different resources available can be integrated into classroom activities. One reason for this is that many of the tools explored were not originally developed for pronunciation training, so teachers may not immediately recognize them as potentially useful for this purpose. Another key aspect is that studies addressing the use of technology for pronunciation practice are typically labeled under the umbrella term "CAPT" (Computer Assisted Pronunciation Training), which may limit teachers' and researchers' understanding of the available tools, especially given that many modern devices no longer require computers in the traditional sense. Moreover, it is common in scientific articles to refer to the potential of "CAPT" or "technology" in general, treating these terms as abstract entities that can enhance the teaching-learning process regardless of what is used or how it is used. Usually, studies inform us of the potential of a specific tool or technique for a specific purpose and under very specific conditions. Thus, general statements about technological affordances may be misleading, as the advantages a given resource may offer do not necessarily apply to other tools.

Fouz-González, J. (2025). Teaching and learning pronunciation with technology: Current possibilities, pedagogical recommendations, and directions for the future. In J. M. Levis, M. Duris, S. Sonsaat-Hegelheimer, & I. Na (Eds.), *Proceedings of the 15th Pronunciation in Second Language Learning and Teaching Conference* (pp. 1-15). Iowa State University, September 2024. https://doi.org/10.31274/psllt.19027

Finally, due to space limitations in research articles, discussing the role of a given technique within the broader framework of pronunciation instruction can be challenging. Hence, even if readers learn that a particular technique or tool is effective, how or when to use it may not be self-evident.

To make the most of the possibilities that technology offers for pronunciation practice, it is important to establish realistic expectations regarding what technology can and cannot do. In this regard, reflecting on some of the concerns researchers have raised over time can be helpful. For example, 26 years ago Pennington (1999) criticized that most computer-based work focused on "the decontextualized mechanics of articulation," had "no curriculum or a limited curriculum," and that "few applications of […] technology link[ed] mechanical and meaningful dimensions of phonology" (p. 432). Twenty years later, O'Brien (2019) raised similar concerns, stating that "current CAPT software does not provide […] many opportunities for true communication" and that "many of the CAPT resources that exist focus only on the development of pronunciation skills and not on the integral connection between perception and production" (p. 337). CAPT applications have also been criticized for their lack of innovation, both pedagogically and technologically (Levis, 2007, p. 185), frequently resembling audiolingual approaches that were mostly focused on repetition, mimicry, and drilling (Rogerson-Revell, 2021, p. 191).

The fact that these concerns continue to be voiced after decades of research is discouraging, especially considering recent technological advancements. Nonetheless, these issues could be approached from a different perspective by reconsidering the role of technology in the teaching-learning process. For example, there is currently no resource that encompasses all the types of practice (e.g., perception and production) that different learners (with different L1s, proficiency levels, and other individual differences) need, in the amount that they need, and tailored to contents that are meaningful to them. But is it realistic to hope for something like this? Is it even necessary? These are expectations that are not typically set for non-technological materials. For instance, when choosing a book for a given course, teachers do not always use it from start to finish, or even in a sequential order. It is common to choose activities that align with the needs of a specific group of learners at a given point in time, but also to adapt them and combine them with other books and materials. Thus, rather than seeking a 'perfect' tool that may never exist, it may be more appropriate to explore the various possibilities available and understand what each tool offers, adopting different resources for different purposes depending on the moment and the needs of each group (Fouz-González, 2015; forthcoming).

As a case in point, consider High Variability Phonetic Training (HVPT), one of the most popular techniques in L2 pronunciation research circles. HVPT is characterized by a focus on very specific features, and training involves bombarding learners with multiple, decontextualized instances of those features (e.g., in words or nonwords). This technique typically targets a single skill (perception), it involves a great deal of repetition—reminiscent of audiolingual drill-like activities—and it originated in the 1990s, so it cannot be considered innovative. Research has consistently shown that it can help learners improve their perception of the L2—one of the biggest challenges they face—even after very short interventions (see Thomson, 2018; Uchihara et al., 2024). HVPT cannot solve all the problems learners have with pronunciation, but that does not mean that it is not useful. It is only *one* of the tools teachers can use in class, and this example of HVPT applies to the use of technology in general.

Since the effectiveness of any technological implementation depends on the pedagogical approach behind it, for an effective integration of technology into the teaching and learning of pronunciation, it is not only important to be aware of what each tool allows us to do (which will be reviewed in the following section), but also to consider what learners need at different stages of the teaching-learning process (discussed in the Linking Technology and Pedagogy section).

ENHANCING PRONUNCIATION TEACHING AND LEARNING WITH TECHNOLOGY

There are multiple ways in which technology can support the teaching and learning of pronunciation. To showcase the potential of the tools available, the next sections start with a brief reflection on how different types of practice could be implemented without any technological help, followed by a summary of the research available on how technology can enhance these types of practice.

Enhancing Perception Training

Some technologies are so "normalized" (Bax, 2003) in language teaching—and particularly in pronunciation instruction—that it is hard to imagine training without any technological support. Non-technological perception training would probably involve similar activities (e.g., identification or discrimination), such as:

a) What do you hear? *live – leave*,
b) Listen to the following text and underline:
   i. instances of /iː/ (as in *sheep*) and /ɪ/ (as in *ship*)
   ii. stressed syllables
   iii. weak forms
c) Which word is different? *cat – cat – cut.*

If no technology were used (not even the traditional audio examples found in supporting materials accompanying paper-based books), the only input/model learners would have would be their teacher. However, with technology, perception practice can be enhanced in multiple ways.

As mentioned above, the most well-known perception training technique nowadays is HVPT, which exposes learners to multiple instances of the target features, pronounced by multiple speakers, in various phonetic contexts. Studies have shown that HVPT can promote robust gains in less than four hours of training (e.g., Carlet & Cebrian, 2022) and that the perceptual gains achieved after only four 30-minute HVPT sessions are comparable to those fostered by a full year of intensive language instruction (Nagle et al., forthcoming).

Access to input (and therefore perceptual training) can also be facilitated with speech synthesis. Text-to-speech (TTS) synthesis allows teachers and learners to obtain pronunciation samples of any word. Unlike traditional dictionaries in which sample audio files are limited to words in their root form, TTS allows users to create audio samples of different word forms (e.g., past tenses, plurals) in longer stretches of speech, and these samples can be customized (e.g., in terms of the voices used, gender, speed, etc.). Studies have shown that this type of input can foster comparable gains to teacher-based training (e.g., Liakin et al., 2017) and that it can be suitable for creating HVPT activities (Qian et al., 2018).

Researchers have also explored the potential of complementing auditory training with visual information on the acoustic signal to direct learners' attention to the target pronunciation features. This additional, unbiased modality—in that it does not involve L1-based perception—can help learners focus on specific features of the input and notice them. Studies have shown that audiovisual training can be superior to audio-only training (Motohashi-Saigo & Hardison, 2009), although the differences between conditions—even when audiovisual training is superior—have not always been significant (Okuno & Hardison, 2016).

Perception can also be trained by directing learners' attention to pronunciation when exposed to audiovisual materials like films or TV shows. As noted in Fouz-González and Mora (2024), these materials, like HVPT, provide intensive exposure to highly varied input, where the target pronunciation features appear in various phonetic contexts and pronounced by multiple speakers, but in context and in communicative situations. Captions can serve as an additional source of information to help learners process speech (Wisniewska & Mora, 2020). In fact, captions have been found to help learners improve their pronunciation of target words (Mohsen & Mahdi, 2021), and textually-enhanced captions have been found to effectively direct learners' attention to specific pronunciation features (Fouz-González & Mora, 2024) and help update learners' phonolexical representations of target words (Galimberti et al., 2023).

Enhancing Production Training

Without any technological assistance, production practice would typically involve individual (e.g., repeating after the teacher or reading aloud) or pair work (e.g., dialogues) face to face. Feedback would be limited to teacher feedback (conditioned by the number of students in class and the time available), peer feedback, and learners' own monitoring of their production, which would not be ideal considering their L1-based perception bias. Nevertheless, with technology, a wide range of possibilities becomes available.

Merely recording learners' production—now possible in almost any device—is a considerable improvement, as this allows capturing speech material that is otherwise transient and difficult to analyze. However, nowadays there are apps with more sophisticated capabilities, such as the MTSR app used in Foote and McDonough's (2017) study on shadowing. This app not only allows users to record themselves while listening to a model—ideal for shadowing—but it also offers enhanced analysis capabilities, enabling users to easily listen to the recordings, individually or together. In Foote and McDonough's study, learners improved their ability to imitate a model, their comprehensibility, and their fluency after practicing four times a week for at least ten minutes over six weeks.

Another way in which technology can enhance production practice is through tools that not only facilitate the recording but also the sharing of audios and subsequent feedback provision, such as podcasting services, voice blogs, and social networking sites (SNSs). For example, podcast-like activities in which learners record and share their pronunciation to receive feedback have been found to foster gains in global (Lord, 2008) and specific pronunciation features (Fouz-González, 2019). Likewise, studies using voice blogs (Hsu, 2016) and SNSs (Sun et al., 2017) have shown that having a real audience motivated students and encouraged them to work harder and revise their submissions more.

Production training can also be enhanced with visual information about pronunciation. As in perception practice, visual information can support the auditory modality by directing learners' attention to specific pronunciation features that might otherwise be overlooked. In production practice, visual information can enhance both models for imitation and analyses of learners' pronunciation. CAPT research has addressed visualizations of acoustic or articulatory information. Studies on acoustic visualizations have shown the potential of pitch contour displays to learn intonation (e.g., Ramírez-Verdugo, 2006), prosody (e.g., Hardison, 2004), and Mandarin tones (e.g., Chun et al., 2015); of waveform and spectrographic displays to teach specific segmental features (e.g., Olson, 2014); and of plots of vowel formant values to train vowel sounds (e.g., Rehman, 2021). Studies on visualizations of articulatory information have shown the potential of techniques such as ultrasound tongue imaging (Roon et al., 2020), electropalatography (Hacking et al., 2017), electromagnetic articulography (Suemitsu et al., 2015), and talking heads (Peng et al., 2018) to help learners improve their articulation of challenging segments. These articulatory visualization techniques require specialized equipment that may be challenging to incorporate into language classes. However, recent studies have addressed other innovative and more accessible possibilities, like the digital zoom technology of smartphones' front cameras to amplify the articulators (Wang et al., 2024) or Augmented Reality (AR) filter technology in mobile apps to highlight the user's facial features that are relevant for sound articulation (Zhu et al., 2024).

Studies show that production practice can also be assisted with speech resynthesis, a technique in which learners typically pronounce utterances and receive feedback in the form of a recast with their own voice (adjusting the acoustic parameters of mispronunciations to align with those of a native speaker). The rationale behind the technique is that imitating a similar voice should facilitate noticing potential differences between the two productions (Probst et al., 2002) and that hearing a better version of themselves (Henderson & Skarnitzl, 2022) may have a motivating effect by making the task seem more attainable. Studies have shown that resynthesis can help learners notice their mistakes and improve their pronunciation in different ways, including for lexical stress (Bissiri & Pfitzinger, 2009), intonation (De Meo et al., 2013), fluency, and comprehensibility (Ding et al., 2019). Nevertheless, speech resynthesis has not always been found to be superior to natural speech (De Meo et al., 2013), with some studies showing that natural speech with emphasis on the target features can be more effective (Bissiri & Pfitzinger, 2009).

Since learners' self-monitoring relies on their perception of their pronunciation—likely biased by their L1—a great deal of research has addressed tools that evaluate speech automatically. Studies have investigated software incorporating automatic error detection—specifically designed to offer feedback on pronunciation—and dictation tools that use automatic speech recognition (ASR) technology. Both technologies offer learners an external and objective evaluation of pronunciation, unlimited attempts in a location of their choosing—thereby ensuring the privacy many learners appreciate—and access at any time. Studies have shown that this type of practice is perceived as less intimidating than interpersonal exchanges (Mroz, 2018) and that both techniques can help learners notice their mistakes and improve their pronunciation of specific segments (e.g., Neri et al., 2008 for error detection; McCrocklin, 2019 for ASR-based dictation) and global pronunciation measures (e.g., Eskenazi et al., 2007 for error detection; Evers & Chen, 2022 for ASR-based dictation). Even though ASR-based practice has traditionally involved controlled activities, typically in the form of read-aloud tasks, current AI-powered chatbots such as Gemini allow learners to engage in spontaneous conversations while viewing a real-time transcript of what the system understands.

Finally, technology can also facilitate production practice by providing learners with more opportunities for interaction. For example, videoconferencing software and virtual worlds allow learners to easily connect with speakers from around the world. Moreover, in the absence of human interlocutors, chatbots and intelligent personal assistants can also serve as conversational partners that are always available. In this regard, research has shown that interpersonal exchanges through videoconferencing software (Bueno-Alastuey, 2010) and interactions with intelligent personal assistants can help learners improve their pronunciation (Tai, 2022), and that the exposure to the language virtual worlds offer can foster gains in the pronunciation of target words (Yamazaki, 2018).

Enhancing Awareness-Raising Activities and the Delivery of Metalinguistic Information

While research suggests that providing learners with explicit information about pronunciation plays an important role in L2 pronunciation development (Couper, 2011; Saito, 2013), CAPT studies have mostly focused on the potential of technology to assist perception and production training. Technology can also enhance activities intended to raise learners' awareness of how certain pronunciation features work as well as the delivery of metalinguistic information about pronunciation. Examples of these activities include illustrating how the target features are pronounced, comparing similar L2 sounds to one another and to other L1 sounds, or providing information about the target aspects (e.g., common spellings, when to use a particular intonation contour, or how a sound is articulated and how it differs from those of the L1). Without technology, this would be done in class, with the teacher simply illustrating these aspects or presenting information orally. With technology, most of the techniques described in the preceding sections could be used for these purposes at an initial stage in which the targets are introduced before engaging in more extensive perception or production practice.

For example, Baran-Łucarz et al. (2015) evaluated the potential of clickers to facilitate the delivery of phonetic information. Learner response systems can make classes more interactive with large groups or with students who are reluctant to speak. Anonymous responses to teacher's questions encourage participation and provide teachers with real-time feedback on student understanding. In Baran-Łucarz et al.'s study, clickers fostered engagement and learners were less inhibited to respond. Nevertheless, they were only superior to traditional teaching for one of the target aspects taught.

Technology can also enrich the delivery of pronunciation information by providing it with a special time and space. Studies have addressed the potential of different tools to send learners explicit tips about challenging words and audiovisual examples illustrating them, showing that this approach can help learners improve their pronunciation of the target words and that gains are maintained over time (e.g., Fouz-González, 2017 with Twitter; Saran et al., 2009 with multimedia messages; Xodabande, 2017 with Telegram).

Finally, tools like YouGlish can also be quite useful for this type of activity, as YouGlish provides examples of the target pronunciations in context, pronounced by multiple speakers, and in many contexts. As a case in point, Sardegna and Jarosz (2022) used YouGlish for perception practice following instruction on stress placement in academic words and found that it helped learners improve their perception of lexical stress.

LINKING TECHNOLOGY AND PEDAGOGY

The preceding sections offer an overview of some of the ways in which technology can enhance pronunciation training. However, knowing when to use each tool or whether technology is necessary at all is not always self-evident. For an effective integration of technology into pronunciation instruction, it would be important to first reflect on the types of pronunciation activities required for a given course and then consider the role that different technological tools and techniques could play to support those types of practice.

Pronunciation teaching frameworks typically recommend starting instruction with awareness-raising activities (including descriptions), followed by listening discrimination exercises (though some group both together), and then moving on to production practice, in which activities gradually progress from controlled to spontaneous production (e.g., Celce-Murcia et al., 2010; Zielinski & Yates, 2014; Baker, 2021). To evaluate how technology could enhance these stages, it is helpful to consider each activity type separately: (1) distinguishing between awareness-raising and listening discrimination (as suggested in Celce-Murcia et al., 2010), (2) considering activities with varying degrees of control in both perception and production practice, and (3) addressing a type of activity that is rarely used in the field, focused on consolidating declarative knowledge of pronunciation (for a detailed account, see Fouz-González, forthcoming). Thus, activities could be divided into those aiming at:

1. Raising awareness
2. Consolidating declarative knowledge
3. Perception training
4. Production practice and monitoring

Activities for raising awareness include those that illustrate the pronunciation of the target features or provide learners with explicit information (e.g., articulatory information, sound-spelling correspondences, and common learner difficulties). To raise learners' awareness of a given pronunciation target, teachers could use tools like YouGlish, learner response systems, social media, speech synthesis, or visual information (acoustic or articulatory). To share explicit information about pronunciation, for example, they can employ learner response systems or social media.

Activities aimed at consolidating learners' declarative knowledge include, among others, activities designed to help learners remember sound-spelling correspondences (e.g., "Underline all instances of /æ/ in the following dialogue"), the phonological composition of words (e.g., "Is 'sugar' pronounced with /s/ or with /ʃ/?"), or pronunciation-related rules (e.g., "Which words are stressed/unstressed in this dialogue?") These activities could be used at different stages of the teaching-learning process. For instance, before perception training, to help learners remember useful information about the pronunciation targets that may boost the effectiveness of perceptual training, or after it, to help learners remember or review the phonological composition of words after having trained their perception of certain sounds or patterns. This is important because learners may improve their perception of two challenging sounds after training (e.g., the /i/-/ɪ/ contrast for Spanish EFL learners), but that does not necessarily mean that they know when to use each—especially in English, given the complex nature of sound-spelling relations. Unless learners are exposed to the target words after having learned to perceive the contrast, they are unlikely to know how those words are pronounced. This type of activity is intended to help learners consolidate their pronunciation knowledge, which should in turn be helpful for production practice (helping them know when to use the targets and monitor their pronunciation). Activities like these could easily be created as interactive quizzes using software like Google Forms, Kahoot, etc., or in experimental design software like TP or Gorilla.

Perception training activities are typically implemented with a strong focus on forms and limited context (e.g., HVPT) to help learners notice how the targets are pronounced and to attune their perception to the cues that are necessary for the accurate perception of the L2. When introducing an aspect that is particularly challenging for the first time, highly controlled activities are necessary to help learners focus on pronunciation form (e.g., with little context, using nonwords or short words, like HVPT). However, once learners can perceive the target features in such conditions, activities could extend the context in which those features are presented. For example, this could involve activities in which learners listen to longer stretches of speech (and therefore cannot focus exclusively on form) while drawing their attention to pronunciation (e.g., Fouz-González, 2019; Fouz-González & Mora, 2024; Wisniewska & Mora, 2020). Perception training activities can be easily created with experimental software like Gorilla, but also with simpler tools, such as PowerPoint (e.g., Martin, 2020), Kahoot, and Google Forms. Stimuli for such training can be obtained from platforms like YouGlish, podcasts, audiovisual materials like TV shows, or even generated with TTS (Qian et al., 2018).

Production training in CAPT research has predominantly focused on relatively controlled activities in which learners pronounce words or sentences in isolation and receive immediate feedback. This may be necessary when dealing with challenging pronunciation features for the first time, as controlled activities facilitate drawing learners' attention to the target features. The studies reviewed above have shown that this type of practice can be supported with recordings, visual feedback (acoustic and articulatory), and software that evaluates learners' pronunciation automatically. Nevertheless, technology can also enhance less controlled practice. For example, the analysis of spontaneous interactions in class can be enhanced with tools like VEO (see e.g., Fouz-González & Sánchez-Muñoz, 2020), which allows users to tag and annotate specific moments of the conversation, rate various pronunciation aspects while recording, and then easily go back to further analyze these moments. Learners can also keep a podcasting channel or voice blog in which they share oral tasks (which may be more or less focused on meaning/form) with peers and/or the teacher for feedback. Similarly, spontaneous production can be practiced with tools like Salukispeech (McCrocklin et al., 2022), which requires learners to describe pictures on the spot and provides immediate ASR-based feedback, or with AI-powered chatbots like ChatGPT or Gemini, which allow users to engage in spontaneous conversations and view the transcript of what the system understands. Finally, technology can also support completely spontaneous practice by enabling learners to engage in interactions through oral exchanges (e.g., via videoconferencing software like Skype or Zoom, or in virtual worlds or virtual reality). Even if these exchanges are extemporaneous, they can still have specific pronunciation goals if the teacher encourages learners to monitor their production of features previously covered in class.

THE FUTURE OF CAPT

What Lies Ahead

As new tools or functionalities within existing tools emerge, new possibilities will arise to facilitate different aspects of the teaching and learning of pronunciation. Considering recent advances in artificial intelligence (AI), it is reasonable to expect substantial improvements in many of the possibilities mentioned above (e.g., more natural-sounding TTS, more precise error detection, and more efficient ways to resynthesize speech). Nevertheless, as stated in the introduction, it is unlikely that a single tool can provide learners with all the types of practice they need in the amount that they require. Thus, teachers will always play a central role in the teaching-learning process because they have a unique ability to identify individual challenges and adapt instruction accordingly (Lima & Wallace, 2024).

Challenges to Be Addressed

The disconnect between technology and pedagogy is a perennial problem in CAPT. Even though this concern has been repeatedly highlighted by researchers (Pennington, 1999; Levis, 2007; Derwing & Munro, 2015; Rogerson-Revell, 2021), there are still applications and courseware that are not grounded in research and that have limited pedagogical value. This may be partly due to the absence of quality control requirements for software releases (Kaiser, 2018; Rogerson-Revell, 2021). Introducing an external evaluation process by pronunciation specialists before making educational software available could help address this problem. Alternatively, pronunciation researchers could conduct more software reviews, and journals could make these freely available through open access. These reviews could even be linked to the software's download page, serving as a form of quality assurance to inform users of the software's potential usefulness and to provide recommendations for its effective implementation. Finally, more collaborations between pronunciation specialists and software developers would also help address this issue (Derwing & Munro, 2015; O'Brien et al., 2018).

Another challenge is the limited lifespan of some commercial software. For instance, when mobile applications gained popularity, publishers released promising pronunciation apps based on existing materials developed by pronunciation experts, such as *Clear Speech from the Start* (Cambridge University Press, 2011), *English File Pronunciation* (Oxford University Press, 2012), and *Sounds: The Pronunciation App* (Macmillan Publishers Ltd., 2011). The potential of some apps was even tested empirically (Fouz-González, 2020). Nonetheless, updates failed to keep pace with operating system upgrades, often rendering the apps unusable for weeks or months. Eventually, all these apps were discontinued, leaving users without a product they had paid for—something that would not happen with other types of materials, such as textbooks.

Directions for Future Research

Limited participant numbers and the absence of control groups are common issues in many CAPT studies. Therefore, in addition to exploring the potential of new tools and functionalities, the field would benefit from replication studies that address these limitations.

Additionally, many CAPT studies have focused on ways to improve controlled production practice. Future research should also investigate tools to enhance spontaneous practice, thereby expanding the range of tools available to help learners move from controlled to spontaneous pronunciation knowledge.

Also, most CAPT studies concentrate on enhancing perception and production training. Nevertheless, given the role explicit information plays in L2 pronunciation development, it would also be important to explore ways in which technology can facilitate the delivery of metalinguistic information and help learners consolidate their declarative pronunciation knowledge.

Finally, since research typically investigates the effectiveness of tools and techniques under specific—and controlled—conditions, the field would benefit from studies that evaluate how often and for how long learners use technology on their own (see e.g., Walesiak, 2021), as well as qualitative studies exploring how they use technology independently, their views and experiences, and the impact of more autonomous uses of technology on learners' pronunciation.

Strengthening the Link between Research and Practice

As noted above, even if we learn that a given tool or technique can positively impact learners' pronunciation, knowing how or when to use it may not be self-evident. To strengthen the connection between research findings and classroom practices, it is essential that studies provide a more detailed account of the training procedures and pedagogical implications. If discussing pedagogical implications in detail is not possible due to the journal's requirements, authors could link research articles to less technical publications more focused on pedagogy (e.g., the Teaching Perspectives available in the *Journal of Second Language Pronunciation* or articles in the *Speak Out!* journal published through IATEFL's Pronunciation Special Interest Group).

Finally, while CAPT studies provide information on tools and techniques that are effective, readers may not always be able to use them because these tools have been specifically created for a given study and are not publicly available. Since researchers invest significant time and effort in designing tools and training paradigms, it would be important to make these resources accessible to the public. This would be particularly helpful for teachers who wish to incorporate technology into their classes but do not know how to create the resources themselves. Examples of research tools that have been made available for general use include the *English Accent Coach* (Thomson, 2012-2025), which offers extensive HVPT training; the *Golden Speaker Builder* (Ding et al., 2019), which provides controlled production practice and feedback through speech resynthesis (Ding et al., 2019); *iSpraak* (Nickolai, 2015), which offers controlled production practice with ASR-based feedback; and Salukispeech (McCrocklin et al., 2022), which provides opportunities for spontaneous practice with ASR-based feedback.

ABOUT THE AUTHOR

Jonás Fouz-González is an Associate Professor in the Department of Didactics of Language and Literature at the University of Murcia (Spain), where he teaches courses in English phonetics and pronunciation teaching as well as several EFL modules. His research interests include computer-assisted pronunciation training, L2 speech learning and pronunciation teaching, and English phonetics and phonology. Most of his research to date has focused on the ways in which technology can facilitate pronunciation training.

Contact information:
jfouz@um.es

REFERENCES

Baker, A. (2021). 'She'll be right': Development of a coaching model to clear and fluent pronunciation in Australia. *English Australia Journal*, *37*(1), 27–39.

Baran-Łucarz, M., Czajka, E., & Cardoso, W. (2015). Teaching English phonetics with a learner response system. In E. Waniek-Klimczak & M. Pawlak (Eds.), *Teaching and researching the pronunciation of English* (pp. 35–61). Springer. https://doi.org/10.1007/978-3-319-11092-9_3

Bax, S. (2003). CALL–past, present and future. *System, 31*(1), 13–28. https://doi.org/10.1016/S0346-251X(02)00071-4

Bissiri, M. P., & Pfitzinger, H. R. (2009). Italian speakers learn lexical stress of German morphologically complex words. *Speech Communication*, *51*(10), 933–947. https://doi.org/10.1016/j.specom.2009.03.001

Bueno Alastuey, M. C. (2010). Synchronous-voice computer-mediated communication: Effects on pronunciation. *CALICO Journal*, *28*(1), 1–20. https://doi.org/10.11139/cj.28.1.1-20

Cambridge University Press. (2011). *Clear speech from the start.* [iPhone app].

Carlet, A., & Cebrian, J. (2022). The roles of task, segment type, and attention in L2 perceptual training. *Applied Psycholinguistics*, *43*(2), 271–299. https://doi.org/10.1017/S0142716421000515

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). *Teaching pronunciation. A reference for teachers of English to speakers of other languages, 2nd ed.* Cambridge University Press.

Chun, D. M., Jiang, Y., Meyr, J., & Yang, R. (2015). Acquisition of L2 Mandarin Chinese tones with learner-created tone visualizations. *Journal of Second Language Pronunciation*, *1*(1), 86–114. https://doi.org/10.1075/jslp.1.1.04chu

Couper, G. (2011). What makes pronunciation teaching work? Testing for the effect of two variables: Socially constructed metalanguage and critical listening. *Language Awareness, 20*(3), 159–182. https://doi.org/10.1080/09658416.2011.570347

De Meo, A., Vitale, M., Pettorino, M., Cutugno, F., & Origlia, A. (2013). Imitation/self-imitation in computer-assisted prosody training for Chinese learners of L2 Italian. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th pronunciation in second language learning and teaching conference* (pp. 90–100). Iowa State University.

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

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

Eskenazi, M., Kennedy, A., Ketchum, C., Olszewski, R., & Pelton, G. (2007). The NativeAccent pronunciation tutor: Measuring success in the real world. *Proceedings of Speech and Language Technology in Education (SLaTE2007)*, 124–127.

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

Foote, J., & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation, 3*(1), 34–56. https://doi.org/10.1075/jslp.3.1.02foo

Fouz-González, J. (2015). Trends and directions in Computer Assisted Pronunciation Training. In J. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 314–342). Palgrave Macmillan.

Fouz-González, J. (2017). Pronunciation instruction through Twitter: The case of commonly mispronounced words. *Computer Assisted Language Learning*, *30*(7), 631–663. https://doi.org/10.1080/09588221.2017.1340309

Fouz-González, J. (2019). Podcast-based pronunciation training: Enhancing FL learners' perception and production of fossilised segmental features. *ReCALL*, *31*(2), 150–169. https://doi.org/10.1017/S0958344018000174

Fouz-González, J. (2020). Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. *Language Learning & Technology*, *24*(1), 62–85. https://doi.org/10125/44709

Fouz-González, J. (forthcoming). *Technology-assisted pronunciation training: Bridging research and pedagogy*. University of Toronto Press.

Fouz-González, J. & Mora, J. C. (2024). Exploring the potential of textually-enhanced captioned video to direct learners' attention to challenging sound contrasts. *Journal of Second Language Pronunciation, 10*(3), 309-338. https://doi.org/10.1075/jslp.24043.fou

Fouz-González, J., & Sánchez-Muñoz, A. R. (2020). El etiquetado de vídeo como facilitador del análisis de la pronunciación de la lengua extranjera. *Docencia e Investigación*, *31*, 24–49.

Galimberti, V., Mora, J. C., & Gilabert, R. (2023). Audio-synchronized textual enhancement in foreign language pronunciation learning from videos. *System*, *116*, 103078. https://doi.org/10.1016/j.system.2023.103078

Hacking, J. F., Smith, B. L., & Johnson, E. M. (2017). Utilizing electropalatography to train palatalized versus unpalatalized consonant productions by native speakers of American English learning Russian. *Journal of Second Language Pronunciation, 3*(1), 9–33. https://doi.org/10.1075/jslp.3.1.01hac

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology, 8*(1), 34–52. http://dx.doi.org/10125/25228

Henderson, A. J., & Skarnitzl, R. (2022). "A better me": Using acoustically modified learner voices as models. *Language Learning & Technology, 26*(1), 1–21. http://hdl.handle.net/10125/73462

Hsu, H-.C. (2016). Voice blogging and L2 speaking performance. *Computer Assisted Language Learning, 29*(5), 968–983. https://doi.org/10.1080/09588221.2015.1113185
Kaiser, D. (2018). Mobile-assisted pronunciation training: The iPhone pronunciation app project. *IATEFL Pronunciation Special Interest Group Journal*, *58*, 38–52.
Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27*, 184–202. https://doi.org/10.1017/S0267190508070098
Liakin, D., Cardoso, W., & Liakina, N. (2017). The pedagogical use of mobile speech synthesis (TTS): Focus on French liaison. *Computer Assisted Language Learning*, *30*(3-4), 348–365. http://dx.doi.org/10.1080/09588221.2017.1312463
Lima, E. & Wallace, L. (2024). Friend or foe: Considering what pronunciation teachers can do that ChatGPT cannot. In D. J. Olson, J. L. Sturm, O. Dmitrieva, & J. M. Levis (eds), *Proceedings of 14th Pronunciation in Second Language Learning and Teaching Conference*, (pp. 1-11). Iowa State University.
Lord, G. (2008). Podcasting communities and second language pronunciation. *Foreign Language Annals, 41*(2), 364–379. https://doi.org/10.1111/j.1944-9720.2008.tb03297.x
Macmillan Publishers Ltd. (2011). *Sounds: The pronunciation app.* [iPhone app].
Martin, I. A. (2020). Pronunciation development and instruction in distance language learning. *Language Learning & Technology*, *24*(1), 86–106. https://doi.org/10125/44711
McCrocklin, S. (2019). 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., Fettig, C, & Markus, S. (2022). Salukispeech: integrating a new ASR tool into students' English pronunciation practice. In J. Levis & A. Guskaroska (eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference* (pp. 1–10). Iowa State University. https://doi.org/10.31274/psllt.13342
Mohsen, M.A., & Mahdi, H.S. (2021). Partial versus full captioning mode to improve L2 vocabulary acquisition in a mobile-assisted language learning setting: Words pronunciation domain. *Journal of Computing in Higher Education, 33,* 524–543 (2021). https://doi.org/10.1007/s12528-021-09276-0
Motohashi-Saigo, M. & Hardison, D. M. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Technology, 13*(2), 29–47. http://dx.doi.org/10125/44179
Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals, 51*(3), 617–637. https://doi.org/10.1111/flan.12348
Nagle, C., Mompean, J. A., & Fouz-González, J. (forthcoming). Does timing matter? Exploring different windows of maximal opportunity to enhance the effectiveness of high variability phonetic training.
Neri, A., Cucchiarini, C., & Strik, H. (2008). The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. *ReCALL*, *20*(2), 225–243. https://doi.org/10.1017/S0958344008000724
Nickolai, D. (2015). *iSpraak.* [Computer software]. https://www.ispraak.net
O'Brien, M. G. (2019). Targeting pronunciation (and perception) with technology. In N. Arnold & L. Ducate (Eds.), *Engaging language learners through CALL* (pp. 309–352). Equinox Publishing Ltd.

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., & Muller Levis, G. (2018). 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

Okuno, T. & Hardison, D. M. (2016). Perception-production link in L2 Japanese vowel duration: Training with technology. *Language Learning & Technology*, *20*(2), 61–80. https://dx.doi.org/10125/44461

Olson, D. (2014). The benefits of visual feedback on segmental production in L2 classrooms. *Language Learning & Technology, 18*(3), 173–192. http://dx.doi.org/10125/44389

Oxford University Press. (2012). *English File Pronunciation.* [iPhone app].

Peng, X., Chen, H., Wang, L., & Wang, H. (2018). Evaluating a 3-D virtual talking head on pronunciation learning. *International Journal of Human-Computer Studies, 109*, 26–40. https://doi.org/10.1016/j.ijhcs.2017.08.001

Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning, 12*(5), 427–440. https://doi.org/10.1076/call.12.5.427.5693

Probst, K., Ke, Y., & Eskenazi, M., (2002). Enhancing foreign language tutors – In search of the golden speaker. *Speech Communication, 37*(3–4), 161–173. https://doi.org/10.1016/S0167-6393(01)00009-7

Qian, M., Chukharev-Hudilainen, E., & Levis, J. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *Language Learning & Technology*, *22*(1), 69–96. https://dx.doi.org/10125/44582

Ramírez Verdugo, D. (2006). A study of intonation awareness and learning in non-native speakers of English. *Language Awareness, 15*(3), 141–159. https://doi.org/10.2167/la404.0

Rehman, I. (2021). *Real-time formant extraction for second language vowel production training.* [Unpublished doctoral dissertation]. Iowa State University. https://dr.lib.iastate.edu/handle/20.500.12876/PrMBy1pz

Rogerson-Revell, P. M. (2021). Computer-assisted pronunciation training (CAPT): Current issues and future directions. *RELC Journal, 52*(1), 189–205. https://doi.org/10.1177/0033688220977406

Roon, K. D., Kang, J., & Whalen, D. H. (2020). Effects of ultrasound familiarization on production and perception of nonnative contrasts. *Phonetica, 77*(5), 350–393. https://doi.org/10.1159/000505298

Saito, K. (2013). Re-examining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition, 35*(1), 1–29. https://doi.org/10.1017/S0272263112000666

Saran, M., Seferoglu, G., & Cagıltay, K. (2009). Mobile assisted language learning: English pronunciation at learners' fingertips. *Eurasian Journal of Educational Research, 8*(34), 97–114. http://hdl.handle.net/20.500.12416/944

Sardegna, V. G., & Jarosz, A. (2022). Exploring how *YouGlish* supports learning English word stress: A perception study. In V. G. Sardegna & A. Jarosz (Eds.), *Theoretical and practical developments in English speech assessment, research, and training* (pp. 165–184). Springer. https://doi.org/10.1007/978-3-030-98218-8_10

Suemitsu, A., Dang, J., Ito, T., & Tiede, M. (2015). A real-time articulatory visual feedback approach with target presentation for second language pronunciation

learning. *The Journal of the Acoustical Society of America*, *138*(4), EL382–EL387. https://doi.org/10.1121/1.4931827
Sun, Z., Lin, C.-H., You, J., Shen, H. J., Qi, S., & Luo, L. (2017). Improving the English-speaking skills of young learners through mobile social networking. *Computer Assisted Language Learning, 30*(3–4), 304–324. https://doi.org/10.1080/09588221.2017.1308384
Tai, T.-Y. (2022) Effects of intelligent personal assistants on EFL learners' oral proficiency outside the classroom. *Computer Assisted Language Learning*, *37,* (5–6), 1281–1310. https://doi.org/10.1080/09588221.2022.2075013
Thomson, R. I. (2018). High Variability [Pronunciation] Training (HVPT). A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208–231. https://doi.org/10.1075/jslp.17038.tho
Thomson, R. I. (2012-2025). *English Accent Coach.* [Computer software]. https://www.englishaccentcoach.com
Uchihara, T., Karas, M., & Thomson, R. I. (2024). Does perceptual high variability phonetic training improve L2 speech production? A meta-analysis of perception-production connection. *Applied Psycholinguistics*, *45*(4), 591–623. https://doi.org/10.1017/S0142716424000195
Walesiak, B. (2021). Mobile apps for pronunciation training: Exploring learner engagement and retention. In A. Henderson, A. Kirkova-Naskova, & J. Fouz-González (Eds.), *English pronunciation instruction: Research-based insights* (pp. 357–384). John Benjamins. https://doi.org/10.1075/aals.19.15wal
Wang, S., Li, J., & Liang, Q. (2024). Visual reinforcement through digital zoom technology in FL pronunciation instruction. *Language Learning & Technology*, *28*(1), 1–26. https://hdl.handle.net/10125/73558
Wisniewska, N. & Mora, J. C. (2020). Can captioned video benefit second language pronunciation? *Studies in Second Language Acquisition, 42*(3), 599–624. https://doi.org/10.1017/S0272263120000029
Xodabande, I. (2017). The effectiveness of social media network telegram in teaching English language pronunciation to Iranian EFL learners. *Cogent Education*, *4*(1), 1347081. https://doi.org/10.1080/2331186X.2017.1347081
Yamazaki, K. (2018). Computer-assisted learning of communication (CALC): A case study of Japanese learning in a 3D virtual world. *ReCALL*, *30*(2), 214–231. https://doi.org/10.1017/S0958344017000350
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
Zielinski, B., & Yates, L. (2014). Myth 2: Pronunciation instruction is not appropriate for beginning-level learners. In L. Grant (Ed.), *Pronunciation myths: Applying second language research to classroom teaching* (pp. 56-79). University of Michigan Press.

[1] The PSLLT 2024 plenary talk and this short article are based on the contents of my forthcoming book *Technology-assisted pronunciation training: Bridging research and pedagogy* (University of Toronto Press). Please note that this is a highly condensed version. Readers interested in a more in-depth discussion are encouraged to consult the book.