12
TECHNOLOGY AND PRONUNCIATION

Mª Camino Bueno-Alastuey & Esther Gómez Lacebex

Introduction

In the last few years, the evolution of technological resources has been rapid and has affected all realms of Second Language Acquisition (SLA), including pronunciation. Although L2 pronunciation remained an underexplored area until the turn of the 21st century and is still neglected in many EFL contexts, it is now enjoying renewed interest. Recent review articles (Lee et al., 2015; Thomson & Derwing, 2015), and handbooks (Kang et al., 2018; Reed & Levis, 2015) have updated research carried out so far, and identified state-of-the-art on-going and future lines of research. Undoubtedly, the availability of technologies has made it possible not only to increase and refine research on pronunciation, but also to improve pronunciation teaching.

The study of speech has exploited mechanical, electronic, digital, and, more recently, ultrasound and magnetic resonance devices. Early ones such as kymographs, which were used to register articulatory and phonation details tracing sound waves on a rotating cylinder; or gramophones, used to reproduce and record sound, were expensive and normally limited to the phonetics laboratory and to those few who studied foreign languages. The advent of the magnetic tape, representative of the well-known audio-lingual interest for pronunciation and of the so-called 'language laboratories,' made 'listen and record' more affordable and accessible for the growing number of foreign language learners in the 1960s. Phoneticians and phonologists were able to store and analyze more audio data, which they could also complement with the articulatory information provided by still photography and video tape. Nevertheless, the biggest change took place with computerization and digitalization as the job of instruments such as oscillographs (record air pressure), spectrographs, cameras, and tape recorders could be performed by a single cheap, accurate, and portable machine. Nowadays, technology is able to provide acoustic measurements of speech (speech analysis), generate speech (speech synthesis), and identify speech (speech recognition). These three main components of speech technology have significantly contributed to the recent increasing interest in the study of second language speech and pronunciation, along with an increasing compromise for collaboration between foreign language teachers and second language researchers, and the redemption of the notion of intelligibility (Levis, 2005) which is now seen as the targetable goal in pronunciation teaching. In fact, pronunciation learning and technology soon allied in computer-assisted pronunciation teaching (CAPT), a fruitful research field which explores the efficacy of technologies for pronunciation teaching and learning with the current challenges of elucidating how we can best use their affordances for the sake of learners' communicative success.

DOI: 10.4324/9781351117586-15

In this chapter, we will first provide an overview of present critical issues related to pronunciation in SLA and how technology can contribute to their development. Second, we will present how technology has influenced research on second language speech. And third, we will give an overview of the main advances in pronunciation teaching and learning that technology has allowed exposing some current limitations and potentials.

Current Research on Critical Issues and Topics

Research on pronunciation has focused on critical issues to understand the process of learning pronunciation such as input, intelligibility, etc. This section will review those critical issues and concerns at present, analyzing how technology has deepened researchers' understanding of them, and has contributed to their development.

Input

The quality and the quantity of the amount of language that a learner is exposed to has been a traditional concern in SLA. The abundant (copious), varied (exposed to different speakers' voices), and authentic (meaning oriented) input that the child, or first language acquirer, receives has been associated with successful language acquisition. However, second language learners in environments where the target language is not copious (e.g., not spoken in the community), or varied (e.g., limited to teacher's voice), and with little chances to interact genuinely for communication have often been predicted to struggle with pronunciation development. Technology may be in the position to redesign both the quantity, and the quality of input by allowing access to abundant and varied speech when learners socially network, access internet, or stream entertainment. For example, social media permits speakers and listeners to engage in genuine conversation that does not meet space and/or time boundary limitations. In addition, although still inceptive and not totally successful (Beneteau et al., 2019), intelligent personal assistants (IPA) have enabled communication between humans and machines (Jurafsky & Martin, 2014). Such new communication exchange contexts are likely to shape the role of input in foreign language acquisition by dissociating it from input-poor conditions. Finally, research has questioned the measurement of input by macrovariables such as age of arrival to the target language community, residing years for immigrants, or number of hours/years receiving instruction for language learners (Flege, 2018). Technology can now accompany individuals at all times, which allows for non-invasive collection of data about linguistic exchanges and enables recalling information for shorter and more recent periods of time, thus, providing more rigorous accounts of the amount and quality of L2 input (Flege, 2009).

Speech Production

Foreign accent in L2 learners' production is mainly interpreted in terms of the influence, or *transfer*, of phonetic and phonological patterns of the L1 onto the L2 (Flege & Bohn, 2021). However, research has also revealed that such influence can decrease with time and that more universal acquisition patterns such as overgeneralization also operate during L2 speech production. Universal phonological complexity (*markedness*) or phonetic or articulatory settings, or the tendency to maintain certain articulatory configurations which can be language-specific, have also been used to account for L2 speech. However, to date, L2 phonological production remains to become comprehensively incorporated in theories of L2 speech (Colantoli et al., 2015; Zampini, 2008). One of the possible reasons for the fact that the speech production domain has not been sufficiently addressed may be that it has been mainly inferred and interpreted via phonetic acoustic instrumentation. Today, technology offers new measuring techniques which can inform us more precisely about the physical and articulatory dimension of speech production such as ultrasound imaging for

tongue surface data (Gick et al., 2008), or electromagnetic articulography (EMA), which provides measures about position and movement inside the mouth, or flesh-point tracking, which provides facial gestural information during speech acts. In addition, neuroimaging techniques such as magnetic resonance imaging (MRI) or electrophisiological techniques such as electroencephalography (EEG) or electrocorticography (ECoG) (Mesgarani et al., 2014) allow for looking into the brain by detecting electrical activity of motor events. Such new instrumentation, however, offers the challenge of still being costly and complex to interpret. In addition, it has not overcome the challenge of normalizing inter-speaker variability and simultaneous measures are still difficult to perform. Nevertheless, there is no doubt that as measuring tools become less invasive and more affordable, we will keep refining acoustic interpretation of articulation and will be able to account for the simultaneity and composition of speech production. Finally, the problem of elicitation in speech analysis has for long been envisaged. L2 speech research is almost always programmed under monitored recording conditions for two main reasons 1) the wish to guarantee the quality of the speech signal, which forces the researcher to use dedicated speech recording equipment, and somewhat unnatural phonetic booths to reduce background noise; and 2) the need to guarantee that the target structure is uttered. A recent call for exploring natural casual or spontaneous speech in SLA has been made (Lozano & Medikoetxea, 2013). Such speech mode may allow for investigating more underlying phonological aspects, speaking styles, suprasegmentality, or connected speech, to mention but some features that may not be well represented in elicited speech. The latest technological developments offer less invasive and more economical recording devices similar in recording qualities to those of more sophisticated equipment. Furthermore, technology has also allowed research to better understand the circumstances of a L2 listener in authentic communication, very likely to happen in noise (García Lecumberri et al., 2010), or to investigate which neural perceptual strategies L2 listeners activate when they process speech in noise (Rammell et al., 2019).

Intelligibility

The global spread of English, its international status, and the number of native and increasing non-native varieties of English that are spoken today is likely to have contributed to the need to justify accent variability, and as a consequence, the focus of pronunciation development has moved from sounding like a native to being intelligible (Lewis, 2005). The concept of intelligibility itself is elusive in its definition, conceptualization, and assessment. Munro and Derwing (1995) distinguished between three different constructs: 'intelligibility,' or the extent to which an utterance is understood by a listener, 'comprehensibility,' or the perceived difficulty with which an utterance can be understood, and 'accentedness,' or how different a pattern of speech sounds is from a reference accent or variety. While it would seem reasonable to believe that these three concepts align, research has shown that they do not necessarily correlate as speakers have been found to be heavily foreign-accented and highly intelligible, for example, (Derwing & Munro, 1997). Considerable research has looked into the factors that promote intelligibility and comprehensibility rather independently, such as vowels and consonants with high functional load (Munro & Derwing, 2006), sentence stress (Hahn, 2004), word stress (Field, 2005), or speech rate (Derwing et al., 2004). More recently, research on comprehensibility has been carried out from a more interactional, and dynamic perspective and has focused on how linguistic (Saito et al., 2016) and non-linguistic aspects such as listener attitude or perceived anxiety (Kennedy & Trofimovich, 2019) contribute to comprehensibility, or showing that comprehensibility ratings can fluctuate in the same listening task (Nagle et al., 2019). Within these dynamic interpretations of communicability and in order to further our understanding of how different factors affect accentedness, comprehensibility, and intelligibility, technology offers the possibility to integrate several measures within the same study, for example, by monitoring audio and visual interactions, physical data, logging anxiety (Mora &

Mora-Plaza, 2019), or gestures (Zheng & Samuel, 2019) as well as interpreting interaction practices such as repetition, paraphrasing, clarification requests, or accommodation strategies (Kaur, 2017). Among these technologies, Automatic Speech Recognition (ASR) can be used not only to measure the importance of different factors which can contribute to and might explain speakers' intelligibility, but also to enhance learners' awareness of their own intelligibility (Mroz, 2018), even though some researchers have pointed to the superiority of human raters for ratings of intelligibility (Derwing, 2018).

Variability and Individual Differences

SLA research has a long tradition of acknowledging differences in language learning on account of the several factors that may interact in the learning process, which may be different (or maybe unique?) in each individual. Aspects such as age, language use, amount of experience, amount and quality of input, influence of the L1, learning context, and learning strategies have been widely explored in SLA pronunciation studies (Dörnyei, 2006). This line of research, which has traditionally tended to explore one variable at a time, has come to conclude that pronunciation development is shaped by a multitude of factors (Trofimovich et al., 2014), and that these can be interdependent, the study of variability having been referred to as of "obscure academic interest" (Chapelle, 2004), and still not sufficiently addressed in SLA research (Sawyer & Ranta, 2001). Some of these variables such as L1, target language background, and experience need to be incorporated as controlling variables in studies that address groups of speakers, so that the sample is as homogeneous as possible. Technology offers extraordinary data collection opportunities that may be able to capture these variables and incorporate them as including/excluding variables in study designs. For example, Iverson & Evans (2009) took into account the L1 of their participants and success with High Variability Pronunciation Training (HVPT). Nevertheless, individual variability in results has usually been explained by cognitive and psychologically oriented learner-internal factors such as motivation, extraversion, aptitude, personal identity, analytical reasoning, or phonological memory without having taken them into account as variables in the experiment. Technology can also aid in finding more research participants using online calls for participation in order to find subject profiles, which match more closely study variables such as exceptional speakers, motivational profiles, or personality traits. Sociolinguistic research could benefit from computerization with massive data collection, using crowdsourcing tools, or on-line questionnaires, and analyses of data qualitatively and quantitatively, resulting in more powerful statistical calculations. Finally, we must consider the increasing integration of technology in communication and education to have become an active agent in the learning process (Chapelle, 2007). Technology is reshaping language learners' exposure to input and is facilitating new learning tools as learners have been found to compensate for working memory capacity while using a multimedia package in Chun and Payne (2004), for example.

Relationship Between Speech Perception and Production

Research conducted on the interaction between L2 speech perception and production has provoked diverse results as for how both skills develop in the L2 learner and how they influence one another (Levy & Law, 2010; Saito & van Poeteren, 2018). Until recently, attempts at studying such links had been comparisons or correlations of behavioral methods for each aspect: identification/discrimination of (synthesized/natural/variable) perceived speech scores, and auditory/acoustic or holistic analyses of elicited productions. Studies that have attempted to correlate these different variables have provided mixed results to date, not clearly being able to elucidate the assumption that perception leads to production (as in the L1) finding weak correlation indices or non-synchronic development. One way of accomplishing these limitations may be by incorporating the same

measuring technique to both dimensions using neuroimaging techniques such as magnetic resonance imaging (MRI), electrophysiological techniques such as electroencephalography (EEG), or electrocorticography (ECoG) (Mesgarani et al., 2014). These new procedures can also add further knowledge about the subphonemic mental representations of L2 speech, speech perception, and speech processing. Research should also be able to interpret new behavioral responses to linguistic stimuli by measuring reaction times or interpreting the movements of the eyeball (eye-tracking) (Cutler et al., 2006). As for production, new measures such as ultrasound, EMA, or gesture/face tracking can add further detail about articulatory and gestural aspects of speech production. The challenge that researchers face with the multiple speech analysis tools available is to integrate them in research so that different measurements can be sampled (acoustic-auditory, articulatory-acoustic, or behavioral and neural as, for example, in Sereno & Wang, 2007.)

L2 speech training has also been used as a moderating variable in the inspection of the relationship between perception and production (Bradlow, 2008; Sakai & Moorman, 2018; Thomson & Derwing, 2015). However, many of these training studies owed their design to the premise that speech perception is a precursor to speech production. As a consequence, a considerable number of training studies have explored the effects of perception training on production skills (see meta study by Sakai & Moorman, 2018), while few studies (Kartushina et al., 2015) have dealt with the effects of speech production training on L2 perception skills.

Research Methods

Technology has supported and opened new avenues for second language speech research from participant recruitment phases to dissemination stages. The following section will provide a review of the main empirical findings that second language perception and production research has produced by considering how technology has contributed to such results.

Speech synthesis, or the manipulation of the speech signal, has illustrated speakers' progress in the incorporation of the acoustic parameters needed to decode phonological information from language-general to language-specific features at early stages of first language acquisition (Jusczyk, 2000; Werker & Curtin, 2005). As for L2, studies rigorously controlling or manipulating acoustic parameters in their design have shown that language background and language experience can affect L2 vowel (Cebrian, 2006) or consonant (Akahane -Yamada, 1995) perception. Other studies such as Escudero and Boersma (2004) have also shown that L2 learners can accommodate their perceptual learning strategies to different dialects using synthetic speech in their method. However, despite such computer-generated contributions, synthetic speech has been alleged to still sound somewhat artificial to explore perceptual abilities which incorporate the full range of acoustic parameters that are involved in phonetic discrimination and phonological decoding. Hence, in study designs that do not intend to control for specific acoustic cues, researchers can use natural speech material. A good example is the use of HVPT in L2 phonetic training (Pisoni & Lively, 1995), a training paradigm which has been largely used to explore the development of new sounds by L2 speakers (Thomson, 2018). These studies base their design on the premise that exposure to variability of talkers (and other parameters) contributes more effectively to the formation of a new mental representation (or category) of an L2 sound. The stimuli design in these studies involves the recruitment of different speakers, who are asked to elicit speech as required in each experiment.

Computerization and digitalization also play a leading role here by affording researchers the access to speech databases. Although too broadly catalogued (gender and age of the speaker), non-native speech databases such as AMI, The Speech Accent Archive, The Arabic Learner Corpus, and The Spanish Learner Oral Corpus offer the possibility to access previously recorded non-native speech. Also, crowdsourcing tools such as Amazon Mechanical Turk are now being explored as resources for inexpensive and fast data collection and analysis (Nagle, 2019).

Technology also supports speech researchers in stimuli design allowing for high quality voice-recording devices with free software such as Audacity, Anvil, or Praat (Hardison, 2018). Specific software also allows for stimuli processing such as selecting an optimal recording frequency range, reducing/eliminating background noise, or normalizing (averages volume level) for perception tasks, or splicing (cutting) selected material out of the sound file, to mention but a few, maybe most frequent, speech sound processing operations. Specific software allows for annotation (labelling or coding) of sound files for later analysis and also for alignment as in Okuno and Hardison (2016), who coordinated waveform and video information to explore different training modes. In addition, freeware such as PsychoPy or OpenSesame, or software, such as E-Prime or Superlab, permit exporting laboratory actions on-line. These packages offer design and collect interfaces as well as analysis tools for behavioral data. More specific platforms such as TP 3.1 (Rauber et al., 2012) help in the presentation of stimuli for identification and discrimination perceptual testing.

Psycholinguistics is also now exploring non-behavioral methodologies as technological improvements have granted more affordable and portable ultrasound equipment (Gick et al., 2008), which permit closer articulatory inspection of the tongue, for example. Eye-tracking technology is also being exploited in L2 language processing (Mora & Cerviño-Povedano, 2019) and neuroimaging techniques, which localize brain activity during language tasks, and interpret on-line processing (Sabourin, 2009), are also becoming more accessible for linguists.

Technology for L2 Pronunciation from a SLA Perspective

In addition to widening researchers' understanding of speech processes, technological advancements can become extremely useful for teaching and learning pronunciation. Phonetic training studies have contributed to pronunciation acquisition research by elucidating that L2 speakers can create new L2 perceptual categories, improving speech reception and production, after phonetic training (Saito & Plonsky, 2019).

Computer Assisted Pronunciation Training (CAPT) can ameliorate pronunciation acquisition processes and can help overcome some of the main constraints attributed to pronunciation learning in foreign language learning contexts, namely, insufficient input, lack of practice time (insufficient output) and lack of individualized attention and/or feedback. In such circumstances, technology has been suggested as "an avenue for solution" (Fouz-González, 2015, p. 314). CAPT has the capacity to provide high amounts of meaningful and varied input offering virtually unlimited access to input, both in its simplest form providing sound files of individual sounds and words in online dictionaries and exemplifying how any utterance is pronounced using synthetic speech. Through networking sites and websites, CAPT gives access to audio speech and exposure to different varieties of English using multimodal material, overcoming space and time-boundary limitations. Furthermore, pronunciation can be illustrated with more engaging and informative visual and response interfaces, for example, using ultrasound (Alsabaan & Ramsay, 2014; Kocjančič Antolík et al., 2019), HVPT (Thomson, 2018), or adapting phonetic training for younger learners (Gómez-Lacabex & Gallardo-del-Puerto, 2014). Such new language supply may well compensate for the lack of richness, variety, and authenticity that has been accounted for the L2 language learner. Second, CAPT can provide additional speaking practice and more time on task. As the EFL classroom is unable to offer enough opportunities for speaking practice (Ahn & Lee, 2016), CAPT systems can offer real audiences. For example, telecollaboration using technology such as Skype widens the range of interlocutors available and can test intelligibility more accurately and individually. Telecollaboration between students in distant locations with different L1 can contribute to focusing students' attention on errors in their pronunciation because they hinder intelligibility, forcing them to produce pushed output to correct those errors, and improving their pronunciation as the highest rate of uptake in such exchanges takes place on phonological aspects (Bueno-Alastuey, 2010). Furthermore, technology allows interactions to be recorded very easily, which can refrain students from using their

L1 as they feel the teacher may listen to the recordings, or even make it unfeasible in the case of using telecollaboration with interlocutors who do not know students' L1 (Bueno-Alastuey, 2013).

Social networking apps such as Twitter have also been explored with positive outcomes to increase time on explicit pronunciation instruction of problematic aspects of the foreign language (Fouz-González, 2017). Research is also ready to explore the potential of already existing technologies such as the incorporation of multimedia through film extracts (Wisniewska, & Mora, 2020), radio interviews, or podcasts (Ducate & Lomicka, 2009; Fouz-González, 2019); or the use of 3D animations (Alsabaan & Ramsay, 2014); or the adoption of a game-based approach with simulations and role-plays, and their effect on pronunciation development (Tejedor-García et al., 2020). As an example, Lys (2013) used iPads and reported an increase in the amount and quality of oral production of the students because they spent a considerable amount of time recording their own speech and engaging in activities that facilitated interactions outside the classroom. This increase in practice time led the students to become more comfortable and competent about their speaking. Increased opportunities for interaction, forcing students to reformulate, self-correct, and persist trying to make themselves understood by Artificial Intelligence technologies (i.e., IPAs such as Alexa) have also been reported (Underwood, 2017; Moussalli & Cardoso, 2020). Third, experts (e.g., Neri et al., 2002) have highlighted the relevance of providing effective feedback, addressing specific difficulties, and suggesting remedy for pronunciation improvement. Technology can improve significantly the provision of this type of feedback as it enhances the comprehensibility and the reliability of feedback by including visual representations, which have proved to be beneficial for production (Olson, 2014). Visual acoustic and articulatory interfaces can be displayed by computers in the form of waveforms, spectrograms, and articulatory visualizations (Inceoglu, & Gnevsheva, 2020). While waveforms and spectrograms require training for interpretation and may be only marginally useful for teachers and learners, articulatory visualizations such as talking heads (Massaro et al., 2008) are able to provide both a "role-play dialogue system for conversation training" (Wik & Hjalmarson, 2009, p. 1025) and visual representations of the mouth movements needed to produce sounds accurately compared to the articulations which learners have made (Inceoglu & Gnevsheva, 2020). These visualizations appear to give learners the opportunity to judge the degree of accuracy of their productions, as opposed to solely relying on auditory impressions. The benefits of audio and visuals (talking heads) versus audio-only training (i.e., two channels of input vs. one) have been shown to improve the perception and production of nonnative sounds for learners of English (Hardison, 2003; Hazan et al., 2005), and for learners of French (Inceoglu, 2016).

Further technologies useful for pronunciation feedback are those of speech-to-text by means of Automatic Speech Recognition (ASR). They have the potential to give students more objective feedback on the errors of their speech by reproducing in writing what the system identifies, which aids pronunciation acquisition by increasing learners' awareness of their errors or deviations (McCrocklin, 2019). Nevertheless, instructors need to be cautious regarding the program chosen as comparisons between some of them (Google Voice Typing vs Windows Speech Recognition) have reported a great variability in rates of recognition of nonnative speech, specially in free speech conditions, ranging from 60% to 90 % (McCrocklin et al., 2019). This enormous variation in the rate of recognition might cause frustration in students.

Finally, technology, with its increasing mobility and portability can extend the possibilities of practicing anywhere and at any time (Liakin et al., 2015), while also promoting more private and stress-free learning environments, thus, helping to diminish the anxiety factor students might feel when they have to produce language (Kralova Skorvagova et al., 2017). These environments might lower the affective filter (Krashen, 1988) and improve motivation and pronunciation acquisition. Indeed, motivation has been believed to interact with L2 pronunciation development, although the scarce research conducted to date cannot conclude on the existence of a direct correlation between a high motivational profile and high pronunciation achievement (Trofimovich et al., 2014). Technology does offer a potential new pronunciation learning condition and a more individualized learning experience in which the learner may be able to select type of materials, choose the practice

environment, and be able to self-pace. Undoubtedly, however, these more individual learning preferences need monitoring, guidance and feedback, for which pronunciation tutors and specialists are required.

Future Directions

L2 speech and pronunciation research is currently enjoying renewed interest and technology has been a triggering force. Promising research lines are likely to enlighten the impact of input in new communication channels, and the role of production in the development of L2 speech as new measuring techniques such as neuroimaging refine the interpretation of the speech signal. Massive data access and their exchange will further inform researchers on second language learning profiles and preferences, and further specify which features of speech contribute the most to intelligibility using ASR-based measures (O'Brien et al., 2018). Technology now irremediably mediates the language learning process as learners are exposed to the target language via computers and mobile devices as well as teachers implementing CALL in their practices. CAPT is incorporating more resources such as mobile apps, social media, and ASR-based devices to the classroom. However, CAPT is also now in need to validate its pedagogical value (Lee et al., 2015; Mahdi & Al Khateeb, 2019; Thomson & Derwing, 2015).

Training effects reported in reception and production-based studies should be further explored using larger samples, testing the validity of pronunciation improvements from training in a restricted context (i.e., carrier utterance) to more cognitively challenging and naturalistic speech for longer periods, and with more delayed post-tests (Thomson, 2018). More research is also needed at lower levels of proficiency and focusing on both segmental and suprasegmental features. Furthermore, experts insist on research based on the collaboration among classroom specialists, pronunciation researchers, and computer engineers (O'Brien et al., 2018).

Further research should also focus on the creation of an accessible and fully annotated interactional and spoken corpus with a high quantity of varied speakers so that it includes sufficient amounts of representative and reliable data to reveal which errors affect intelligibility and comprehensibility the most (O'Brien et al., 2018). This corpus will allow for the provision of better feedback, and for data to research what makes other non-English L2 speech difficult to comprehend (Thomson & Derwing, 2015). It would also increase the accuracy of ASR for nonnative speech and could contribute to ASR implementation more broadly for open-ended tasks.

We shall not conclude without mentioning two challenges that CAPT faces today. First, the reality of English as a global language and its increasing pronunciation variety presents a pedagogical as well as a technical challenge, as practitioners are now asked to avoid native varieties as a pronunciation reference and to take into consideration intelligibility and comprehensibility constructs, as aiming at those aims is more realistic. In such a context, CAPT should help practitioners to perceive pronunciation teaching and learning as accent addition instead of accent reduction. Second, technology has also provoked the emergence of electronically-mediated communication; computer-mediated speech has been found to be faster, shorter, and exhibiting more disengagement (Ware, 2005), although speaking to a camera has been reported to be perceived as less natural and less comfortable (Kern, 2014). We must remember that intelligibility is indeed affected by the speech transmission channel and it remains to be seen whether such wired interface may have its own contribution to language development as Chun et al. (2016) remind us that communicating tools might change communication. The impact such practices might have on language learning remains unknown.

Further Reading

O'Brien, M. G., Derwing, T. M., Cucchiarini, C., Hardison, D. M., Mixdorff, H., Thomson, R. I., & Levis, G. M. (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

A good review on the role of technology in state-of-the-art pronunciation research and instruction with concrete suggestions for future developments.

Pennington, M. C., & Rogerson-Revell, P. (2019). Using Technology for Pronunciation Teaching, Learning, and Assessment. In *English Pronunciation Teaching and Research* (pp. 235–286). Palgrave Macmillan.
A revision of the main technologies which are being used for pronunciation development more focused on teaching and assessment than on research, but also including the main research advancements technology has allowed. It includes some alternative uses of well-known and very common technologies such as twitter, voki or Edmodo and how to use them to teach pronunciation.

Kang, O., Thomson, R., & Murphy, J. (Eds.). (2018). *The Routledge handbook of contemporary English pronunciation*. Routledge.
Offers an overview of pronunciation from theoretical perspectives, to a description of English phonetics and its varieties as a global language. It reviews the scope of pronunciation teaching and offers two sections with state-of-the-art pronunciation issues and future directions.

Colantoni, L., Steele, J., Escudero, P., & Neyra, P. R. E. (2015). *Second language speech*. Cambridge University Press.
A thorough review of major second language speech topics and experimental research in the form of a research manual to help anyone to start conducting research on the acquisition of second language speech.

Wayland, R. (Ed.). (2021). *Second language speech learning: Theoretical and empirical progress*. Cambridge University Press.
A collection of chapters on the latest core issues on pronunciation research.

References

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

Akahane -Yamada, R. (1995). Age and acquisition of second language speech sounds perception of American English/r/and/l/by native speakers of Japanese. In W. Strange. (Ed.). *Speech perception and linguistic experience: Issues in cross-language research* (pp. 305–320). York Press.

Alsabaan, M., & Ramsay, A. (2014). Diagnostic CALL tool for Arabic learners. In S. Jager, L. Bradley, E. J. Meima, & S. Thouësny (Eds.) *CALL design: Principles and practice, Proceedings of the 2014 EUROCALL Conference* (pp. 6–11). Research-publishing.net.

Audacity Team (2019). *Audacity(R): Free Audio Editor and Recorder* [Computer application]. Version 2.3.1

Beneteau, E., Richards, O. K., Zhang, M., Kientz, J. A., Yip, J., & Hiniker, A. (2019, May). Communication breakdowns between families and Alexa. In Hu, K., Demiralp, Ã., Gaikwad, S. N. S., Hulsebos, M., Bakker, M. A., Zgraggen, E., ... & Satyanarayan, A. (Eds). *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). https://doi.org/10.1145/3290605.3300473

Boersma, P. & Weenink, D. (2019). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.50 www.praat.org/

Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English. In J. G. H. Edwards, & M. L. Zampini (Eds.), *Phonology and second language acquisition* (Vol. 36) (pp. 287–308). John Benjamins.

Bueno-Alastuey, M. C. (2010). Synchronous-voice computer-mediated communication: Effects on pronunciation. *CALICO Journal*, *28*(1), 1–20.

Bueno-Alastuey, M. C. (2013). Interactional feedback in synchronous voice-based computer mediated communication: Effect of dyad. *System*, *41*(3), 543–559. https://doi.org/10.1016/j.system.2013.05.005

Colantoli, L., Steele, J., & Escudero, P. (2015) *Second language speech*. Cambridge University Press.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*(3), 372–387. https://doi.org/10.1016/j.wocn.2005.08.003

Chapelle, C. A. (2004). Technology and second language learning: Expanding methods and agendas. *System*, *32*(4), 593–601. https://doi.org/10.1016/j.system.2004.09.014

Chapelle, C. A. (2007). Technology and second language acquisition. *Annual Review of Applied Linguistics*, *27*, 98–114. https://doi.org/10.1017/S0267190508070050

Chun, D. M., & Payne, J. S. (2004). What makes students click: Working memory and look-up behavior. *System*, *32*(4), 481–503. https://doi.org/10.1016/j.system.2004.09.008

Chun, D., Kern, R., & Smith, B. (2016). Technology in language use, language teaching, and language learning. *The Modern Language Journal*, *100*(S1), 64–80. https://doi.org/10.1111/modl.12302

Derwing, T. M. (2018) Putting an accent on the positive: New directions for L2 pronunciation and instruction. Proc. International Symposium on Applied Phonetics (ISAPh 2018), 12–18, https://doi.org/10.21437/ISAPh.2018-3

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *19*(1), 1–16.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, *54*(4), 655–679. https://doi.org/10.1111/j.1467-9922.2004.00282.x

Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review*, *19*(1), 42–68. https://doi.org/10.1075/aila.19.05dor

Ducate, L., & Lomicka, L. (2009). Podcasting: An effective tool for honing language students' pronunciation?. *Language Learning & Technology*, *13*(3), 66–86. http://dx.doi.org/10125/44192

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 551–585.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*(3), 399–423. https://doi.org/10.2307/3588487

Flege, J. E. (2009). *Give input a chance*! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA*. (pp. 175–190). Multilingual *Matters*.

Flege, J. E. (2018). It's input that matters most, not age. *Bilingualism: Language and Cognition*, *21*(5), 919–920. https://doi.org/10.1017/S136672891800010X

Flege, J. E., & Bohn, O. S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress*. (pp. 3–83). Cambridge University Press.

Fouz-González, J. (2015) Trends and directions in computer-assisted pronunciation training. In J. A. Mompean, & J. Fouz-González (Eds.), *Investigating English pronunciation* (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

García Lecumberri, M. L., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11–12), 864–886. https://doi.org/10.1016/j.specom.2010.08.014

Gick, B., Bernhardt, B. M., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In J. Hansen & M. Zampini (Eds.), *Phonology and second language acquisition* (pp. 309–322). John Benjamins.

Gómez-Lacabex, E. G., & Gallardo-del-Puerto, F. (2014). Two phonetic-training procedures for young learners: Investigating instructional effects on perceptual awareness. *Canadian Modern Language Review*, *70*(4), 500–531. https://doi.org/10.3138/cmlr.2324

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, *38*(2), 201–223. https://doi.org/10.2307/3588378

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522. http://doi.org/10.1017/S0142716403000250

Hardison, D. M. (2018). Visualizing the acoustic and gestural beats of emphasis in multimodal discourse: Theoretical and pedagogical implications. *Journal of Second Language Pronunciation*, *4*(2), 231–258. https://doi.org/10.1075/jslp.17006.har

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, *47*(3), 360–378. https://doi.org/10.1016/j.specom.2005.04.007

Inceoglu, S. (2016). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics*, *37*(5), 1175–1199. http://dx.doi.org/10.1017/S0142716415000533

Inceoglu, S., & Gnevsheva, K. (2020). Ultrasound imaging in the foreign language classroom: Outcomes, challenges, and students perceptions. In O. Kang, S. Staples, K. Hirschi, & K. Yaw (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching Conference* (pp.115–126). Northern Arizona University.

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866–877. http://dx.doi.org/10.1121/1.3148196

Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing* (Vol. 3). Pearson.
Jusczyk, P. W. (2000). *The discovery of spoken language*. MIT Press.
Kang, O., Thomson, R., & Murphy, J. (Eds.). (2018). *The Routledge handbook of contemporary English pronunciation*. Routledge.
Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, *138*(2), 817–832. https://doi.org/10.1121/1.4926561
Kaur, J. (2017). Intelligibility in global contexts. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation*. (pp. 542–555). Routledge.
Kennedy, S., & Trofimovich, P. (2019). Comprehensibility: A useful tool to explore listener understanding. *The Canadian Modern Language Review*, *75*(4), 275–284.
Kern, R. (2014). Technology as Pharmakon: The promise and perils of the internet for foreign language education. *The Modern Language Journal*, *98*(1), 340–357. https://doi.org/10.1111/j.1540-4781.2014.12065.x
Kipp, M. (2017). Anvil 6.0. The Video Annotation research Tool. www.anvil-software.org/#
Kocjančič Antolík, T., C. Pillot-Loiseau & T. Kamiyama (2019). The effectiveness of real-time ultrasound visual feedback on tongue movements in L2 pronunciation training. *Journal of Second Language Pronunciation*, *5*(1), 72–97. https://doi.org/10.1075/jslp.16022.ant
Kralova, Z., Skorvagova, E., Tirpakova, A., & Markechova, D. (2017). Reducing student teachers' foreign language pronunciation anxiety through psycho-social training. *System*, *65*, 49–60. https://doi.org/10.1016/j.system.2017.01.001
Krashen, S. D. (1988). *Second language acquisition and second language learning.* Prentice-Hall International.
Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345–366. https://doi.org/10.1093/applin/amu040
Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369–377. https://doi.org/10.2307/3588485
Levy, E. S., & Law, F. F. (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America*, *128*(3), 1290–1305. https://doi.org/10.1121/1.3466879
Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, *32*(1), 1–25.
Lozano, C., & Mendikoetxea, A. (2013). Learner corpora and second language acquisition: The design and collection of CEDEL2. In A. Díaz Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 65–100). John Benjamins.
Lys, F. (2013). The development of advanced learner oral proficiency using iPad. *Language Learning & Technology*, 17 (3), 94–116.
Mahdi, H. S., & Al Khateeb, A. A. (2019). The effectiveness of computer-assisted pronunciation training: A meta-analysis. *Review of Education 7*(3), 733–753. https://doi.org/10.1002/rev3.3165
Massaro, D. W., Bigler, S., Chen, T., Perlman, M., & Ouni, S. (2008). Pronunciation training: The role of eye and ear. In *Ninth Annual Conference of the International Speech Communication Association. Interspeech 2008.* (pp. 2623–2626).
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., Humaidan, A., & Edalatishams, E. (2019). ASR dictation program accuracy: Have current programs improved? In *Proceedings of the 10th pronunciation in second language learning and teaching conference* (pp. 191–200).
Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006–1010. DOI: 10.1126/science.1245994
Mora, J. C. & Mora-Plaza, I. (2019). The effect of task complexity and speaking anxiety on L2 fluency, pronunciation accuracy and comprehensibility. Paper presented at the *43 AEDEAN conference*, 13–15 November 2019, Alicante, Spain.
Mora, J. C., & Cerviño-Povedano, E. (2019). The effects of bimodal L2 input on the processing of function words by Spanish EFL learners: An eye-tracking study. In Herrero, C., & Vanderschelden, I. (Eds.), *Using film and media in the language classroom: Reflections on research-led teaching* (pp. 76–91). Multilingual Matters.
Moussalli, S., & Cardoso, W. (2020). Intelligent personal assistants: Can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning*, *33*(8), 865–890. https://doi.org/10.1080/09588221.2019.1595664
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

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. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, *34*(4), 520–531. https://doi.org/10.1016/j.system.2006.09.004

Nagle, C. (2019). Developing and validating a methodology for crowdsourcing L2 speech ratings in Amazon Mechanical Turk. *Journal of Second Language Pronunciation*, *5*(2), 294–323. https://doi.org/10.1075/jslp.18016.nag

Nagle, C., Trofimovich, P., & Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. *Studies in Second Language Acquisition*, *41*(4), 647–672. https://doi.org/10.1017/S0272263119000044

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (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

O'Brien, M. G., Derwing, T. M., Cucchiarini, C., Hardison, D. M., Mixdorff, H., Thomson, R. I., & Levis, G. M. (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. http://llt.msu.edu/issues/june2016/okunohardison.pdf

Olson, D. J. (2014). Benefits of visual feedback on segmental production in the L2 classroom. *Language Learning & Technology*, *18*(3), 173–192. http://llt.msu.edu/issues/october2014/olson.pdf

Pisoni, D. B., & Lively, S. E. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange (Ed.), *Speech perception and linguistic experience. Issues in cross-language research* (pp. 433–462). York Press.

Rammell, C. S., Cheng, H., Pisoni, D. B., & Newman, S. (2019). L2 speech perception in noise: An fMRI study of advanced Spanish learners. *Brain Research*, *1720,* 146316. https://doi.org/10.1016/j.brainres.2019.146316

Rauber, A. S., A. Rato, D. Kluge & G. R. Santos (2012). *TP: Testes / Treina-mentos de Percepção*. www.worken.com.br/tp_regfree.php

Reed, M., & Levis, J. (2015). *The handbook of English pronunciation*. John Wiley & Sons.

Sabourin, L. (2009). Neuroimaging and research into second language acquisition. *Second Language Research*, *25*(1), 5–11. https://doi.org/10.1016/j.brainres.2019.146316

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

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, *37*(2), 217–240. https://doi.org/10.1017/S0142716414000502

Saito, K., & van Poeteren, K. (2018). The perception–production link revisited: The case of Japanese learners' English/ɹ/performance. *International Journal of Applied Linguistics*, *28*(1), 3–17. https://doi.org/10.1111/ijal.12175

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187–224. https://doi.org/10.1017/S0142716417000418

Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 319–353). Ernst Klett Sprachen.

Sereno, J. A., & Wang, Y. (2007). Behavioral and cortical effects of learning a second language. In O-S Bohn & M.J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 239–258). John Benjamins.

Tejedor-García, C., Escudero-Mancebo, D., Cardeñoso-Payo, V., & González-Ferreras, C. (2020). Using challenges to enhance a learning game for pronunciation training of English as a second language. *IEEE Access*, *8*, 74250–74266.

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.

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

Trofimovich, P., McDonough, K., & Foote, J. A. (2014). Interactive alignment of multisyllabic stress patterns in a second language classroom. *TESOL Quarterly*, *48*(4), 815–832.

Underwood, J. (2017). Exploring AI language assistants with primary EFL students. In K. Borthwick, L. Bradley, S. Thouësny (Eds.) *CALL in a climate of change: adapting to turbulent global conditions-short papers from EUROCALL* (pp. 317–321). Research-publishing.net.

Ware, P. (2005). "Missed" communication in online communication: Tensions in a German-American telecollaboration. *Language Learning & Technology*, *9*(2), 64–89.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197–234.

Wik, P., & Hjalmarsson, A. (2009). Embodied conversational agents in computer assisted language learning. *Speech Communication*, *51*(10), 1024–1037.

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

Zampini, M. L., (2008). L2 speech production research. In J. G. H. Edwards, & M. L. Zampini, (Eds.), *Phonology and second language acquisition* (Vol. 36) (pp. 219–249). John Benjamins.

Zheng, Y. & A. G. Samuel (2019). How much do visual cues help listeners in perceiving accented speech? *Applied Psycholinguistics*, *40*(1), 93–109.