Ease and Difficulty in L2 Pronunciation Teaching: A Mini-Review

Mary Grantham O’Brien *

School of Languages, Linguistics, Literatures and Cultures, University of Calgary, Calgary, AB, Canada

Both L2 learners and their teachers are concerned about pronunciation. While an unspoken classroom goal is often native-accented speech (i.e., a spoken variety of the mother tongue that it not geographically confined to a place within a particular country), pronunciation researchers tend to agree that comprehensible speech (i.e., speech that can be easily understood by an interlocutor) is a more realistic goal. A host of studies have demonstrated that certain types of training can result in more comprehensible L2 speech. This contribution considers research on training the perception and production of both segmental (i.e., speech sounds) and suprasegmental features (i.e., stress, rhythm, tone, intonation). Before we can determine whether a given pronunciation feature is easy or difficult to teach and—more importantly—to learn, we must focus on: 1) setting classroom priorities that place comprehensibility of L2 speech at the forefront; and 2) relying upon insights gained through research into L2 pronunciation training. The goal of the mini-review is to help contextualize the papers presented in this collection.

Keywords: second language, pronunciation, training, priorities, effectiveness, comprehensibility

INTRODUCTION

Researchers and teachers alike agree that most adult second language (L2) learners will not sound like native speakers and that speaking with a nonnative accent is normal (Derwing and Munro, 2009). Nonetheless, both teachers and students express a desire for learners to achieve native-accented speech (Timmis, 2002; Sifakis and Sougari, 2005; Scales et al., 2006). Thus, the nativeness principle (i.e., a belief that nativelike pronunciation is both achievable and enviable (Levis, 2005; Levis, 2020)), serves as an implied objective in many language classrooms. In spite of this, recent studies demonstrate that teachers engage only intermittently in classroom pronunciation training, primarily because they lack training (Derwing and Munro, 2015) or confidence (Baker, 2011) or because they have relatively little knowledge about how to teach and assess pronunciation (Baker and Murphy, 2011; Baker, 2014; Couper, 2017). When they do teach pronunciation in their classrooms, teachers tend to focus on segmental production (Foote et al., 2016; Levis, 2016; Couper, 2017), most probably because materials—especially textbooks—tend to focus on segments (Derwing et al., 2012a; Foote et al., 2016). It is not surprising that teachers might be reluctant to teach pronunciation if their ultimate objective is native-accented speech. However, a host of recent studies have demonstrated that being understood is a more realistic goal (Derwing and Munro, 2015). The intelligibility principle, with its acknowledgment that most foreign-accented speech is comprehensible1, thus guides recent L2 pronunciation research (Levis, 2005; Levis, 2020). Researchers generally agree that both segments

1Intelligibility and comprehensibility are terms that are used in research to describe methods for testing listeners’ understanding of speech. Levis’s (2005, 2020) intelligibility principle incorporates both intelligibility and comprehensibility.
and suprasegmental features play an important role in being understood (Derwing and Munro, 2015) and that explicit pronunciation training can have a positive impact on the comprehensibility of L2 speech (Derwing et al., 1998; Isaacs, 2009; Lee et al., 2014; Thomson and Derwing, 2015).

Given the unspoken classroom goal of native-accented speech coupled with the sporadic attention paid to pronunciation on the one hand, and the research focus on comprehensible speech and a recommendation for regular pronunciation instruction on the other hand, there is clearly a disconnect between pedagogical practice and research findings. This contribution’s focus on teaching pronunciation therefore considers the notions of ease and difficulty from two perspectives: 1) setting classroom priorities that place comprehensibility of L2 speech at the forefront; and 2) relying upon insights into research-informed L2 pronunciation training.

DEFINING EASE AND DIFFICULTY IN L2 PRONUNCIATION TEACHING

Determining whether a given pronunciation feature—segmental or suprasegmental—is more or less difficult to learn depends on the extent to which improvement is shown after training. Given the variation in how pronunciation features are trained, how speech samples are elicited (e.g., reading individual words, sentences or paragraphs; repetition of a model speaker; semi-spontaneous or spontaneous utterances), and how improvement is measured (e.g., acoustic analyses, listener intelligibility tasks, listener ratings of comprehensibility and/or foreign accentedness), the field of L2 pronunciation research does not have an agreed-upon standard for determining whether a given type of training is successful. Nonetheless, the results of two recent meta-analyses have shown that pronunciation instruction almost always leads to improvement (Lee et al., 2014; Thomson and Derwing, 2015).

As a starting point in distinguishing between easy and difficult pronunciation features, it is important to consider the factors that may play a role in L2 pronunciation. First among these is language pairings: the combination of a learner’s first language (L1) and their L2. Studies investigating similar groups of L1 learners of the same L2 often report conflicting results. For example, although the Japanese speakers in Haslam (2011) did not show improvement in English /l/ and /u/ production even after training, other studies have shown improvement on these same segments among Japanese learners (e.g., Hardison, 2003; Hazan et al., 2005). The Mandarin native speakers who were trained in English vowel perception in Wang (2002) did not improve in their production of English vowels, but those in Thomson (2011) did. Given these inconsistent findings, it is clear that other factors must be at play in the ultimate success of pronunciation training. As such, L2 pronunciation researchers look beyond language pairings in their assessments of success of a given type of training. Additional factors may include participant’s age of learning (Aoyama et al., 2008; Baker, 2010), quality of target language interactions (Derwing and Munro, 2015), motivational factors (Nagle, 2018), and learners’ involvement in instructional decisions (Jenkins, 2004).

SETTING PRIORITIES

When it comes to determining which pronunciation features are easy and which are hard to learn, some research has shown that certain features are so easy to learn that they do not need to be trained. For example, the Mandarin- and Slavic-speaking learners of English in Derwing et al. (2012b) demonstrated a ability to accurately perceive sentence stress, intonation and the -teen/-ty distinction in the absence of instruction. While we should not deduce from such findings that accurate perception will result in accurate production, it makes little sense to train such features—in this case the perception thereof—in the classroom or to investigate their development. Moreover, individual variation is also quite common, and certain exceptional learners may not require training. For example, two Dutch-speaking learners of Slovak in Hanulíková et al. (2012) demonstrated native-like perception and pronunciation of Slovak consonant clusters after only 15 min of exposure to the language. It is thus important to know which pronunciation features learners have mastered so that teachers do not waste time focusing on features that do not need to be trained.

In order to determine which pronunciation features learners have difficulty with and thus which should be the focus of classroom training, instructors are encouraged to develop a pronunciation needs assessment as described by Derwing and Munro (2015). Instructors should consider collecting both read and extemporaneous speech samples and assessing the samples both globally and analytically to determine learners’ difficulties. The authors note that a perceptual task that requires learners to demonstrate their ability to perceive relevant segmental and suprasegmental distinctions can further guide the development of a pronunciation curriculum.

With the results of an assessment in hand, teachers are able to set priorities for their classrooms. Those pronunciation features that both cause difficulty and affect learners’ comprehensibility—or those with the highest functional load (Catford, 1987)—should be the focus of training. At the segmental level, functional load can be determined, among other things, on the basis of the number of minimal pairs that are distinguished by two segments. For example, contrasting /l/ and /u/ distinguishes more English words than does producing a contrast between /d/ and /θ/ (Munro and Derwing, 2006). Although researchers have not established a functional load hierarchy for prosodic features of English, lexical (Zielinski, 2008; Isaacs and Trofimovich, 2012) and sentential stress assignment (Hahn, 2004) both play an important role in being understood. While we have a good idea of which pronunciation features of

---

1 An anonymous reviewer brought up the important point that it is possible to teach something well and for learners not to learn it. As such, the issue that we are most concerned with is that of learnability.

2 Readers are reminded that L2 learners of English may not require training in the perception of sentential stress assignment as demonstrated by Derwing et al. (2012b).
English play a central role in understanding speech, that work is lacking for other target languages. Thus, when setting both segmental and suprasegmental pronunciation priorities in classes with target languages other than English, teachers are encouraged in their evaluation of their students’ pronunciation needs assessments to consider the extent to which producing given distinctions plays a role in their ability to understand their students’ speech.

EVALUATING THE EFFECTIVENESS OF TRAINING

Language learners—especially those in the early stages of language learning—tend to show improvement in their pronunciation over time. Thus, in order to determine whether a given type of training is effective, it is important when conducting research to include both a comparison group that receives a different type of training and a control group that receives no training. In addition, a delayed posttest allows researchers to determine whether the effects of training are long lasting (Thomson and Derwing, 2015).

Pronunciation improvement can be determined in two main ways: listener ratings and acoustic analyses. While listener ratings of understanding are considered the gold standard in pronunciation research (Derwing and Munro, 2009), some training studies also make use of acoustic analyses. Much of the research investigating the effectiveness of pronunciation training uses measures of understanding including comprehensibility ratings (e.g., Foote and McDonough, 2017; Martin, 2018) or intelligibility tasks (e.g., Derwing et al., 2014), often together with ratings of fluency and/or foreign accentedness. Acoustic analyses, completed by hand (e.g., Counselman, 2015) or automatically (e.g., Suemitsu et al., 2015; Tejedor-García et al., 2020) are also common and can be used to determine the extent to which certain pronunciation features change over time. Researchers note, however, that significant acoustic differences may not align with listener judgments (Derwing and Munro, 2015).

While few classroom teachers are able to carry out systematic analyses of their students’ pronunciation development, they are encouraged to rely upon pronunciation training methods whose effectiveness has been demonstrated via research. Some of this work is outlined below.

RESEARCH-INFORMED PRONUNCIATION TRAINING

After setting priorities, the next step is to choose how to most effectively train pronunciation. While a teacher’s status as a native or nonnative speaker of the target language does not play a role in learners’ ultimate pronunciation (Levis et al., 2016), the results of research have generally demonstrated that explicit, form-focused instruction along with corrective feedback provides the greatest benefits to learners (Saito and Lyster, 2012; Saito, 2013). Derwing et al. (2014) describe an emergent training program designed to meet English language learners’ (L1 = Vietnamese or Khmer) workplace needs. The classroom instruction, which targeted both perception and production, focused on those aspects of the participants’ speech that affected their intelligibility (i.e., consonant clusters, rhythm and intonation). Participants’ comprehensibility improved after only 17 h of classroom-based training.

A relatively large number of recent studies have investigated the effectiveness of ways to train pronunciation outside of the classroom. Researchers point to a number of benefits of computer-assisted pronunciation training (CAPT). These include unlimited practice time and flexibility as well as opportunities for varied input and immediate feedback (Engwall et al., 2004; Levis, 2007). Gao and Hanna (2016) indicate a further benefit: a computer’s capacity for providing “infinite, patient modeling” (p. 214). An element of fun is also often added to CAPT. For example, Barcomb and Cardoso (2020) demonstrate the effectiveness of gamified pronunciation training (i.e., training that includes elements of a game but that is not actually a game). The Japanese junior high school learners of English in that study were rewarded with points and badges as they completed a series of metalinguistic tasks and perception and pronunciation activities focusing on English /l/ and /ɹ/. Listeners in the study demonstrated both increased metalinguistic awareness and improved pronunciation accuracy over time. While a range of CAPT activity types exist, this contribution will focus on three that have been shown to play a positive role in improving learners’ production: 1) listen and repeat; 2) perceptual training; and 3) visualization.

Although the effectiveness of traditional listen and repeat pronunciation tasks may be limited (O’Brien, 2019), a popular and effective way of training pronunciation by listening to a recording and then recording oneself is shadowing. The English learners in Foote and McDonough (2017) completed eight weeks of shadowing tasks in which they immediately repeated and recorded themselves while echoing dialogues from a sitcom as closely as possible. The task encouraged learners to focus on suprasegmental aspects of speech. Listeners rated pre-test, mid-training and post-test extemporaneous recordings for comprehensibility, accentedness and fluency. The authors found that learners had positive attitudes toward the activities and that learners’ comprehensibility and fluency improved over time. A number of additional researchers have demonstrated the effectiveness of shadowing for the development of both segments (Zajac and Rojczyk, 2014) and suprasegmental features (Lima, 2015).

Studies have investigated the efficacy of perceptual training for improving production (e.g., Counselman, 2015; Lee and Lyster, 2016; Sakai and Moorman, 2018). A popular and effective means of improving primarily segmental production through perceptual training is high variability phonetic training (HVPT), which trains listeners’ perception with a relatively large quantity of speech samples that are produced by multiple speakers in a range of phonetic contexts (Thomson, 2018). The results of HVPT studies speak in its favor for the improvement of English vowels by native speakers of Greek (Lengeris, 2018), Mandarin (Thomson, 2011) and French (Iversion et al., 2011), as well as for the improvement of English consonants including English /l/ and /ɹ/ by Japanese speakers (Bradlow et al., 1997) and a number of English consonants by Korean learners (e.g., Huensch and Tremblay, 2015; Lee and Hwang, 2016). An additional type of perceptual training that has shown positive results is the use of
speech synthesis systems (Mixdorff and Munro, 2013). For example, Liakin et al., (2017) found that L2 learners of French who made use of a simple text-to-speech (TTS) app on their mobile devices improved similarly to those learners who engaged in conversational practice with, and received feedback on their pronunciation from, their teachers in their in their production of French liaison. A highly innovative synthesis system that has demonstrated great promise generates a synthetic, native-accented version of a speaker’s own voice (Ding et al., 2019). Participants in the study who made use of this so-called “golden speaker” version of their own voices showed improved comprehensibility and fluency.

Visualization techniques—including the use of acoustic displays (i.e., waveforms, spectrograms, and pitch tracks), ultrasound images that provide feedback on articulatory processes, and talking heads that provide learners with access to facial movements—allow learners to receive real-time visual feedback on productions. Tools used for visualization can include those designed for acoustic analyses such as Praat (Boersma and Weenink, 2020) and Audacity (Audacity Team, 2020) along with software that has been designed specifically to focus on L2 learners’ pronunciation (e.g., Godfroid et al., 2017). At the segmental level, researchers have demonstrated that teaching learners how to interpret formant frequencies may enable them to improve their vowel productions, as demonstrated the native speakers of Japanese learning American English /æ/ in Suemitsu et al. (2015).4 The English–Spanish L2 learners in Olson (2019), Offerman and Olson (2016), and Olson and Offerman (2020) who learned to interpret waveforms and spectrograms showing Spanish voice onset time also showed improvement after instruction. A number of researchers advocate for the use of waveforms and spectrograms for the teaching of suprasegmentals, especially duration and intonation (e.g., Levis, 1999; Hardison, 2004; Chun, 2013). For example, Levis and Pickering (2004) demonstrated the effectiveness of teaching contextualized discourse intonation to L2 learners of English by tracking intonation contours. The L2 Japanese learners in Okuno and Hardison (2016) received either audiovisual training consisting of audio files and waveform displays, audio-only training, or no training on vowel duration in Japanese. While participants in both experimental groups showed improvement and the ability to generalize what they learned to novel stimuli and new voices, participants in the audiovisual group improved their productions more than participants in the audio-only group. Similarly, Motohashi-Saigo and Hardison (2009) demonstrated the effectiveness of visualizations in learning vowel length and singleton/geminate distinctions. Chun et al. (2015) showed that L2 learners of Mandarin who compared the pitch contours of their own tone production with those of native speakers improved in their production of tones.

The type of feedback learners receive plays an important role in the extent of their improvement. Lee and Lyster (2016) investigated the effect of different types of corrective feedback on a series of perceptual tasks on the production accuracy of Korean-English L2 learners’ vowels. Corrective feedback that took the form of either 1) rejection (i.e., indicating that the chosen answer was wrong) together with the target form; or 2) rejection together with the nontarget form was more effective than feedback that included either 3) a rejection along with both the target and nontarget forms; or 4) rejection only. The authors take this as evidence that providing learners with feedback indicating that their responses are incorrect is not sufficient for learning to occur.

It is important to consider that computer software designed to assess pronunciation “is not based on any particular theory or model of pronunciation which differentiates variation from (true) error” (Pennington, 1999; p. 431). As such, most CAPT promotes accuracy over intelligibility (Levis, 2007). Finally, although automatic speech recognition (ASR), which relies on a combination of acoustic analyses and artificial intelligence, has been touted as a promising way to evaluate and provide feedback on pronunciation (O’Brien et al., 2018), a number of researchers point to the relatively few studies that align ASR error detection and human judgments of speech (e.g., Chun, 2013; Chen and Li, 2016; Johnson and Kang, 2017; McCrocklin and Edalatishams, 2020).

### ADDITIONAL FACTORS

In addition to the type of pronunciation training and feedback learners receive, a number of other factors play a role in the success of training. Central among these is learner awareness. Although research has generally shown that learners have difficulty assessing their own pronunciation (e.g., Trofimovich et al., 2016), learners’ awareness of pronunciation features may be positively related to listeners’ comprehensibility ratings of their speech (Kennedy and Trofimovich, 2010). Explicit tasks that encourage awareness may be especially beneficial. For example, Añorga and Benander (2015) demonstrated the effectiveness of tasks that encourage learners to compare their own productions with models. Along similar lines, in addition to carrying out a range of production tasks, the German L2 learners in Martin (2018) completed tasks that required them to distinguish between foreign-accented and native speech. Their comprehensibility improved over time.

Additional factors that may play a role in the effectiveness of pronunciation training can include learners’ proficiency levels, the length of training, and number of trained phonemes (Sakai and Moorman, 2018). Research has demonstrated that learners at lower levels of proficiency tend to make faster progress than more advanced learners (Sakai and Moorman, 2018), that there is an optimal length of pronunciation training (Lee et al., 2014; Olson and Offerman, 2020), and that the number of targeted phonemes should be constrained, possibly to as few as three (Sakai and Moorman, 2018).5

---

5Garcia et al. (2020) demonstrated that the effectiveness of ASR training for the development of some L2 segments.

6Note, however, that Nishi and Kewley-Port (2007) report detrimental effects for training only a subset of vowels or consonants and advocate instead for training the entire set of vowels.
CONCLUSION

Accessing tools to train pronunciation has never been easier. Any language learner has easy access to a multitude of apps that promise to reduce accents quickly and easily. The focus of many of these tools, however, is often highly salient sounds that often do not play a role in comprehensibility and that may never improve after hours of training (Foote and Smith, 2013). This mini-review was written to provide readers of this collection with a background into the field of pronunciation training. Distinguishing between the notions of ease and difficulty in pronunciation teaching is overall much less important than distinguishing between effective and ineffective types of training. This is especially true if we consider the ultimate goal of pronunciation training to be comprehensible L2 speech.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Aforga, A., and Benander, R. (2015). Creating a pronunciation profile of first-year Spanish students. Foreign Lang. Ann. 48 (3), 434–446. doi:10.1111/fla.12151
Aoyama, K., Guion, S., Flege, J. E., Yamada, T., and Akahane-Yamada, R. (2008). The first years in an L2-speaking environment: a comparison of Japanese children and adults learning American English. Int. Rev. Appl. Linguist. 46 (1), 61–90. doi:10.1111/j.1467-9381.2008.00003
Audacity Team (2020). Audacity. Available at: https://www.audacityteam.org/
Baker, A. (2011). Discourse prosody and teachers’ stated beliefs and practices. TESOL J. 2 (5), 263–292. doi:10.5054/tj.2011.259955
Baker, A. (2014). Exploring teachers’ knowledge of second language pronunciation techniques: teacher cognitions, observed classroom practices, and student perceptions. Tesol Q. 48 (1), 136–163. doi:10.1002/tesq.99
Baker, A., and Murphy, J. (2011). Knowledge base of pronunciation teaching: staking out the territory. TESL Can. J. 28 (2), 29–50. doi:10.18806/tesl.v28i2.1071
Baker, W. (2010). Effects of age and experience on the production of English word-final stops by Korean speakers. Biling. Lang. Cognit. 13 (3), 263–278. doi:10.1177/1366728709999986X
Barcomb, M., and Cardoso, W. (2020). Rock or roll? Gamifying an online course management system for pronunciation instruction: focus on English/t/and/l/. CALICO J. 37 (2), 127–147. doi:10.1558/cj.36996
Boersma, P., and Weenink, D. (2020). Praat: Doing phonetics by computer [Computer program] version 6.1.27. Available at: http://www.praat.org/ (Accessed October 25, 2020).
Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (1997). Training Japanese listeners to identify English/r/and/l/: IV. Some effects of perceptual learning on speech production. J. Acoust. Soc. Am. 101 (4), 2299–2310. doi:10.1121/1.418276
Catford, J. C. (1987). “Phonetics and the teaching of pronunciation,” in Current perspectives on pronunciation: practices anchored in theory. Editor J. Morley (Alexandria, VA: Teachers of English to Speakers of Other Languages), 87–100.
Chen, N. F., and Li, H. (2016). “Computer-assisted pronunciation training: from pronunciation scoring towards spoken language learning,” in Asia-Pacific signal and information processing association annual summit and conference (APSIPA), Jeju, South Korea, December 13–16, 2016 (Seoul, South Korea: IEEE), 1–7. doi:10.1109/APSIPA.2016.7820782
Chun, D. M. (2013). “Computer-assisted pronunciation teaching,” in Encyclopedia of applied linguistics. Editor C. A. Chapelle (Oxford, United Kingdom: Wiley-Blackwell), 823–834. doi:10.1002/9781405198431.ueda0172
Chun, D. M., Jiang, Y., Meyer, J., and Yang, R. (2015). Acquisition of L2 Mandarin Chinese tones with learner-created tone visualizations. J. Sec. Lang. Pron. 1 (1), 86–114. doi:10.1080/17519493.2014.943950
Counselman, D. (2015). Directing attention to pronunciation in the second language classroom. Hispania 98 (1), 31–46. doi:10.1353/hpn.2015.0006
Couper, G. (2017). Teacher cognition of pronunciation teaching: teachers’ concerns and issues. Tesol Q. 51 (4), 820–843. doi:10.1002/tesq.354
Derwing, T. M., Di支配brook, L. G., and Foote, J. A. (2012a). How well do general skill ESL textbooks address pronunciation? TESL Can. J. 30 (1), 22–44. doi:10.18806/tesl.v30i1.1124
Derwing, T. M., Thomson, R. L., Foote, J. A., and Munro, M. J. (2012b). A longitudinal study of listening perception in adult learners of English: implications for teachers. Can. Mod. Lang. Rev. 68 (3), 247–266. doi:10.3138/cmlr.1215
Derwing, T. M., and Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. Lang. Teach. 42 (4), 476–490. doi:10.1017/S026144800800551X
Derwing, T. M., and Munro, M. J. (2015). Pronunciation fundamentals: evidence-based perspectives for L2 teaching and research. Amsterdam, Netherlands: John Benjamins.
Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., and Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: a workplace training study. Lang. Learn. 64 (3), 526–548. doi:10.1111/j.112053
Derwing, T. M., Munro, M. J., and Wiebe, G. E. (1998). Evidence in favor of a broad framework for pronunciation instruction. Lang. Learn. 48 (3), 393–410. doi:10.1111/0023-8333.00047
Ding, S., Liberatore, C., Sonašat, S., Lucić, I., Silpachai, A., Zhao, G., et al. (2019). Golden speaker builder: an interactive tool for pronunciation training. Speech Commun. 115, 51–66. doi:10.1016/j.specom.2019.10.005
Engwall, O., Wik, P., Beskow, J., and Granström, B. (2004). Design strategies for a virtual language tutor. 8th international conference on spoken, Jiju Island, South Korea, March 31, 2004 (Seoul, South Korea: ISCA), 1–4. Available at: http://www.speech.kth.se/ctt/publications/papers04/iclplp2004_tutor.pdf.
Foote, J., and McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. J. Sec. Lang. Pron. 3 (1), 34–56. doi:10.1075/jslp.3.1.02foo
Foote, J., and Smith, G. (2013). Is there an app for that? PhD thesis. Montreal, QC, Canada: Concordia University.
Foote, J. A., Trofinovich, P., Collins, L., and Soler Urzúa, F. (2016). Pronunciation teaching practices in communicative second language classes. Lang. Learn. J. 44 (2), 181–196. doi:10.1080/09571736.2013.784345
Gao, Y., and Hanna, B. E. (2016). Exploring optimal pronunciation teaching: Integrating instructional software into intermediate-level EFL classes in china. CALICO J. 33 (2), 201–230. doi:10.1558/cj.v33i2.26054
Garcia, C., Nickolai, D., and Jones, L. (2020). Traditional versus ASR-based pronunciation instruction: an empirical study. CALICO J. 37 (3), 213–232. doi:10.1558/cj.40379
Godfroid, A., Lin, C.-H., and Ryu, C. (2017). Hearing and seeing tone through color: an efficacy study of web-based, multimodal Chinese tone perception training. Lang. Learn. 67 (4), 819–857. doi:10.1111/lang.12246
Hahn, L. D. (2004). Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. Tesol Q. 38 (2), 201–232. doi:10.2307/3588378
Hanulíková, A., Dediu, D., Fang, Z., Basnaková, J., and Huettig, F. (2012). Individual differences in the acquisition of a complex L2 phonology: a study training. Lang. Learn. 62 (2), 79–109. doi:10.1111/j.1167-9922.2012.00707.x
Hardison, D. M. (2003). Acquisition of second-language speech: effects of visual cues, context, and talker variability. Appl. Psycholinguist. 24 (4), 495–522. doi:10.1017/S0142716403000220
Hardison, D. M. (2004). Generalization of computer-assisted prosody training: quantitative and qualitative findings. Lang. Learn. Technol. 8 (1), 34–52. doi:10.125/25228
Thomson, R. I. (2011). Computer-assisted pronunciation training: targeting second language vowel perception improves pronunciation. *CALICO J.* 28 (3), 744–765. doi:10.11139/cj.28.3.744-765

Thomson, R. I., and Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: a narrative review. *Appl. Ling.* 36 (3), 326–344. doi:10.1093/aplin/amu076

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. doi:10.1075/jslp.17038.tho

Timms, I. (2002). Native-speaker norms and international English: a classroom view. *ELT J.* 56 (3), 240–249. doi:10.1093/elt/56.3.240

Trofimovich, P., Issacs, T., Kennedy, S., Saito, K., and Crowther, D. (2016). Flawed self-assessment: investigating self-and other-perception of second language speech. *Bilingualism* 19 (1), 122–140. doi:10.17/S1366728914000832

Wang, X. (2002). *Training Mandarin and Cantonese speakers to identify English vowel contrasts: long-term retention and effects on production.* Burnaby, Canada: Simon Fraser University.

Zając, M., and Rojczyk (2014). Imitation of English vowel duration upon exposure to native and non-native speech. *Poznań Stud. Contemp. Linguis.* 50 (4), 495–514. doi:10.1515/psicl-2014-0025

Zielinski, B. (2008). The listener: No longer the silent partner in reduced intelligibility. *System* 36 (1), 69–84. doi:10.1016/j.system.2007.11.004

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 O’Brien. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.