UC Berkeley
The CATESOL Journal

Title
High Variability Phonetic Training as a Bridge From Research to Practice

Permalink
https://escholarship.org/uc/item/1dg5n862

Journal
The CATESOL Journal, 30(1)

ISSN
1535-0517

Authors
Barriuso, Taylor Anne
Hayes-Harb, Rachel

Publication Date
2018

DOI
10.5070/B5.35970

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

The CATESOL Journal

TAYLOR ANNE BARRIUSO
RACHEL HAYES-HARB
University of Utah, Salt Lake City

High Variability Phonetic Training as a Bridge From Research to Practice

This review of high variability phonetic training (HVPT) research begins by situating HVPT in its historical context and as a methodology for studying second language (L2) pronunciation. Next we identify and discuss issues in HVPT that are of particular relevance to real-world L2 learning and teaching settings, including the generalizability of learning to new situations, in addition to variations in how HVPT is implemented that may promote optimal learning. Primarily, we focus on the relatively limited research that has explored the use of HVPT as a pedagogical tool. We conclude with recommendations for the future regarding the applicability of HVPT to L2 learning and teaching in the real world.

Learning the pronunciation of a new language means learning both to produce and to recognize new speech sounds. The process by which we recognize speech sounds is called *speech perception.* A vast literature has documented the particular difficulty that second language (L2) learners experience when learning to perceive new L2 speech sounds, and there has been a great deal of interest in developing interventions to address this challenge. In the classroom, these interventions will typically rely on speech produced by the teacher or a single voice on language-learning recordings to demonstrate pronunciation features. In the domain of speech research, however, a technique called high variability phonetic training (HVPT) has shown great promise in increasing listeners' ability to perceive nonnative sounds. HVPT uses multiple voices rather than one voice, thus introducing variability into the perception practice. The variability inherent in different voices seems to help L2 learners to perceive new sounds in a more targetlike way.

In this article we review selected HVPT research to understand its relevance to real-world L2 learning and teaching settings. Levis

(2016) noted that in the field of pronunciation, teaching and research often fail to inform each other despite their potential to do so. He highlights HVPT as an "area of research that has great potential to change the way materials are constructed" (p. 425), arguing that it "has great promise in pronunciation training, even on features that seem particularly resistant to instruction" (p. 426). Although full understanding of the suitability of HVPT in language pedagogy requires additional research, language teachers and curriculum designers may benefit from taking advantage of what we already know about this robust line of laboratory and pedagogical research.

A Brief History of HVPT

Speech researchers have long been interested in the role that phonetic variability—the ways that speech sounds differ depending on a variety of factors, including who is producing them and what phonetic contexts they are in—plays in speech processing and learning. After Liberman, Harris, Hoffman, and Griffith (1957) found that (native) speech sounds are perceived as belonging to mental categories (e.g., native English speakers hear all voiceless bilabial stops as /p/ and all voiced bilabial stops as /b/ rather than on a continuum of "voicelessness" to "voicedness"), many new questions arose. These included: (a) whether this ability is innate or learned (Kuhl, 1979); (b) if it is language-specific or perception-general (Eimas, 1974); (c) what factors contribute to category development (Eimas & Corbit, 1973); (d) how flexible these sound-to-category mappings are (e.g., can the boundary of where /p/ ends and /b/ begins be shifted?) once they have been formed (Verbrugge, Strange, Shankweiler, & Edman, 1976); and (e) what the role of phonetic environment on segment identification is (e.g., are L2 learners able to identify that a sound is part of the /ɑ/ vowel category when it is pronounced in a word such as *pop* as readily as when it is pronounced by itself as [ɑ]?) (Strange, Verbrugge, Shankweiler, & Edman, 1976).

Much of this research focused on what was then referred to as the *constancy problem* (Kuhl, 1979, 1983; Shankweiler, Strange, & Verbrugge, 1977). This referred to the somewhat counterintuitive fact that listeners' perception of (native) speech is highly accurate despite the fact that no two productions of the same sound are ever acoustically identical. Through time, speech perception researchers became interested in the consequences of speech variability not only for speech processing for native-speaking adults (see Pisoni, 1979, for a review), but also for infants as well as nonnative-speaking adults, which is also known as *cross-language speech perception* (Perkell & Klatt, 1986).

Research on cross-language speech perception revealed that lis-

teners do not hear phonological contrasts (pairs of sounds that signal meaningful differences such as /p/-/b/ in *pat-bat*) in a nonnative language in the same way (Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973) or with the same accuracy as native listeners (Miyawaki et al., 1975; Werker & Logan, 1985). Initial research probed the development of language-specific speech perception in infancy (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984; see also Best, 1994, for a review) and the difficulty of perceiving nonnative phonological contrasts (Goto, 1971; MacKain, Best, & Strange, 1981; Werker & Tees, 1984). While these early studies focused primarily on characterizing cross-language speech perception, by the 1980s some researchers began asking whether learners could be trained to perceive L2 speech sounds that were difficult for them (Jamieson & Morosan, 1986; McClasky, Pisoni, & Carrell, 1983; Pisoni, Aslin, Perey, & Hennessy, 1982; Strange & Dittman, 1984; Strange & Jenkins, 1978). By the mid-1990s, an important aspect of the difficulty associated with L2 speech perception had been identified: determining which types of acoustic variability are relevant to distinguishing difficult new contrasts (Pisoni, Lively, & Logan, 1994).

In a groundbreaking study, Logan, Lively, and Pisoni (1991) tested whether listening practice involving speech sound variability (i.e., arising from different talkers with different voices and speech patterns) might help learners rule out those differences that are irrelevant to perceiving an L2 sound contrast. It was thought that if listeners were trained to ignore the irrelevant variability, they would then be able to focus on just those acoustic differences that were important to distinguishing the contrasting sounds. To test this hypothesis, Logan et al. recorded natural productions of 68 English minimal pairs differing only in /r/ and /l/ (e.g., *rock-lock*). Each pair was produced by six different talkers. Native Japanese speakers learning English took a pretest, were trained on the contrast, and then took a posttest. In each phase they completed a forced-choice identification task (e.g., Is this word *rock* or *lock*?). The pretest and posttest were identical. The training phase contained 15 training sessions over three weeks. In each session, participants identified the sounds and were given feedback regarding the correct answer after each trial. During each of the 15 sessions, participants heard stimuli from only one talker, listening to five different talkers three times each. A sixth talker's productions were included in the pretest and posttest, but not in the training phase. Results indicated that participants who received training improved in their ability to identify English /r/ and /l/, while those who did not receive training did not show improvement. The amount of improvement varied depending on phonetic environment (where

the sounds occurred in the words). Initial clusters (e.g., *click* vs. *crick*) and between vowels (e.g., *arrive* vs. *alive*) improved more than initial singleton (e.g., *lock* vs. *rock*) and final singleton (e.g., *ball* vs. *bar*) environments, though participants were already very good at identifying the sounds in final position at pretest. The authors concluded that this training procedure "was more robust than earlier training techniques" (1991, p. 874). In other words, it was more successful than any training that had been tried before.

Lively, Logan, and Pisoni (1993) replicated the 1991 study and added a generalization task that included new words produced by both a familiar and an unfamiliar talker. Participants were able to identify new words produced by both talkers, which means that they were able to generalize this new knowledge beyond just the words they were trained on. In a second experiment, participants were trained using words produced by only one talker. These participants showed improvement from pretest to posttest, but they were not able to generalize their learning to the words produced by the new talker. The authors concluded that talker variability aids in robust category formation—that is, the development of an L2 category that is important to the learner in listening and producing the L2.

Lively, Pisoni, Yamada, Tohkura, and Yamada (1994) again used this same procedure in an attempt to replicate earlier results with a new population (monolingual Japanese speakers) and to determine whether the effects of training were retained long term. They found that this new group improved after training, and that participants performed significantly better on new words produced by a familiar talker than on those from an unfamiliar talker. Results of a delayed posttest showed that three months later, participants' performance was still significantly better than at pretest. Six months later, participants' performance was neither significantly greater than pretest nor significantly worse than at posttest. The authors interpreted the lack of a difference between posttest results and the delayed posttest results six months later as suggesting that the high variability training procedure can lead to long-term changes in how the new sound categories are represented in memory.

The high variability training procedure developed by Logan et al. (1991) came to be accepted as an effective laboratory speech-perception training method (Aliaga-García & Mora, 2009). Accordingly, HVPT established itself as a major field of inquiry in the speech sciences. The following discussion highlights research and findings on the topics of HVPT that are of particular relevance to language teaching, especially whether what is learned from HVPT can generalize to new listening contexts.

HVPT and L2 Acquisition

In this section we discuss some methodological variations in HVPT that may be of particular importance to L2 learners and teachers in the real world. Understanding how differences in HVPT studies can result in differences in learning bears directly on whether L2 listening instruction can improve students' pronunciation abilities.

HVPT and Generalizability

Of crucial importance is the question of whether perceptual learning via HVPT results in more targetlike pronunciation. Bradlow, Pisoni, Akahane-Yamada, and Tohkura (1997) studied whether HVPT that did not involve explicit pronunciation training resulted in better pronunciation. In their study, which consisted of 45 HVPT training sessions over three to four weeks, native Japanese speakers completed a pretest and posttest on perception of the English /r/-/l/ contrast using a forced-choice identification task (e.g., Is this *rock* or *lock*?). Their production was also measured via a repetition task. Native English-speaker judges evaluated the production data in two ways. First, they saw the written form of the intended word, and then they heard a single participant's pretest and posttest productions and selected which one was "better." Second, they completed a forced-choice identification task (e.g., Is this *rock* or *lock*?) with participants' pre- and posttest productions. The results showed that native English speakers not only strongly preferred the posttest pronunciations of trained participants, but they were also significantly more accurate in their identification of posttest productions. In a follow-up, Bradlow, Akahane-Yamada, Pisoni, and Tohkura (1999) found that training resulted in perception and production gains, both of which were retained three months later. The authors interpreted these results as "establishing a perception-production link such that successful perceptual learning leads directly to corresponding improvement in speech production" (p. 9). In other words, HVPT listening practice holds promise for helping learners to improve their pronunciation of difficult L2 sounds.

Lambacher, Martens, Kakehi, Marasinghe, and Molholt (2005) investigated whether the production of five English vowels by native Japanese speakers would improve after HVPT. The learners completed a pretest and posttest, during which they read aloud a list of 20 consonant-vowel-consonant (CVC) minimal pairs (e.g., *cop-cup*). Between the tests, learners participated in six 20-minute HVPT sessions over six weeks. Twenty-six native English-speaking judges each listened to a portion of the productions and identified the vowel they heard by clicking on one of five vowels provided on the screen. The productions were also analyzed acoustically. Both the perceptual analysis by the

native speakers and the acoustic analyses indicated that learners' pronunciations of English vowels improved after HVPT, whereas a control group of untrained participants demonstrated no improvement in either analysis. The authors noted that HVPT "can facilitate … production performance by a group of adult L2 learners, even without any explicit production training" (p. 245). Like Bradlow et al. (1997, 1999), this study suggests that work on listening alone shows promise for improving learners' pronunciation.

The promising results obtained by Lambacher et al. (2005) were somewhat tempered by those obtained in Thomson and Derwing (2016). The researchers found that English learners from various first language backgrounds who received 40 sessions of HVPT over a month showed different amounts of improvement in production depending on both the type of stimuli used in the training and the ways that gains were assessed. When the majority of the HVPT was made up of consonant-vowel (CV) stimuli (which were often not real words, such as [ba]), learners demonstrated small productive gains on English vowels in an elicited imitation task; when the HVPT contained mainly real words, learners did not improve. Neither group showed pronunciation gains in a picture-naming task, suggesting that perceptual gains did not generalize to production in more spontaneous speech. The authors suggested that "perceptual training on its own is insufficient to promote maximal improvement" in production (p. 95). From this we can conclude that more research is required to determine the extent to which HVPT can lead to pronunciation gains.

Stimulus Variability

Because variability is a defining characteristic of HVPT, another important question in the context of L2 learning and teaching is what types of variability should be included in the training. Lively et al. (1993) directly investigated two sources of variability: talkers and phonetic environments. In their first experiment, participants were trained with stimuli containing fewer phonetic environments than in Logan et al. (1991), that is, three contexts rather than five. The number of phonetic environments did not affect the size of perceptual gains in trained and untrained environments or with familiar and unfamiliar talkers. In the second experiment, participants were trained with stimuli produced by only one talker. These participants demonstrated perceptual gains in some of the trained phonetic environments, but they did not generalize their improvement to new talkers or phonetic environments. The authors suggested that single-talker training results only in "stimulus-specific learning rather than robust abstract category acquisition" (p. 13). In other words, training that did not

contain multiple talkers helped listeners learn the particular words that they were trained on, but it did not set learners up to develop L2 perceptual categories that they could use in real-world listening.

Improvement May Vary by Learner. Several studies indicate that the effectiveness of HVPT may vary depending on the individual learner. One study (Perrachione, Lee, Ha, & Wong, 2011) involved training by a single talker versus four talkers for Mandarin, a tonal language. The researchers investigated how participants' ability to perceive lexical pitch in Mandarin words depended on their preexisting ability to perceive pitch. They found that participants with higher preexisting pitch-perception ability demonstrated greater learning achievement with the multi-talker training, while those with lower preexisting pitch-perception ability benefited more from single-talker training. For both groups of participants, there was better generalization to new talkers for those who had received multi-talker training than with those who had received only single-talker training.

The authors followed up on these findings by looking more closely at the participants with lower preexisting pitch perception. They used three variations on the multi-talker training procedures and found that when the multi-talker training was blocked by talker (i.e., listeners heard all the productions from talker A, then productions from talker B, and then those from talker C, etc., rather than having all three of the talkers' productions mixed together), these learners showed greater gains. The authors interpreted these results as indicating that the high degree of variability from one word to the next was the source of the relative difficulty of HVPT for the low-sensitivity group. The teaching implication from this research is twofold: (a) Individual differences among learners may affect differences in the effectiveness of HVPT, but (b) these differences might be lessened if we understand the source of the difficulty. Presenting the words talker by talker did not reduce the effectiveness of the multi-talker training for those with high preexisting pitch perception, but it increased the effectiveness of the multi-talker training for those with lower pitch perception. In other words, modifying HVPT methods according to individual learners' abilities may positively affect the effectiveness of HVPT.

The inclusion of multiple talkers and multiple phonetic contexts is inherent to the HVPT paradigm as originally conceived by Logan et al. (1991), and other studies have broadly confirmed these contributions. Nonetheless, studies directly investigating the optimal number of talkers (see e.g., Thomson, 2012c, for an example of a large number of talkers) or the optimal types of phonetic environments have been limited.

Other Training Stimulus Considerations

Researchers have investigated several additional considerations regarding training stimuli that may contribute to the efficacy of HVPT, including how large the training set size is, whether training items (stimuli) are words or nonwords, and whether the training materials are modified or synthetically enhanced.

Training Set Size. Nishi and Kewley-Port (2007) investigated the effect of the number of trained segments on HVPT outcomes. They were interested in two questions: (a) whether listeners are able to learn more than five L2 vowel categories at a time and (b) whether training on a few difficult vowel categories can result in improvement of other categories. Specifically, they investigated whether training including a full L2 vowel set (in this case, American English monophthongs /i, ɪ, ɛ, æ, ɑ, ʌ, ɔ, ʊ, u/) or only the most difficult subset of vowels for L1 Japanese speakers (/ɑ, ʌ, ʊ/) resulted in different perception gains. Participants trained on the subset of vowels showed greater gains on those three vowels than the participants trained on the full set of vowels, but they did not demonstrate any gains on untrained vowels (that is, they were not able to generalize their learning). The full vowel set group demonstrated greater gains overall on a vowel identification task that included the full set of vowels, their gains were better maintained three months later, and they were able to generalize their learning to new talkers and new words better than participants whose training included only a subset of the vowels. The authors concluded that "efficient learning of nonnative vowels requires exposure to the full set of vowel categories, both easy and difficult, in the target language" (p. 1506).

Real Words Versus Nonwords. Another element that has differed among studies is the status of training stimuli as words or nonwords. As noted above, Thomson and Derwing (2016) investigated whether pronunciation gains resulting from HVPT differed depending on whether the majority of training stimuli were nonwords or when training primarily involved real words. Participants' productions of real words in two tasks were recorded before and after training. In the first task, they heard, "The next word is ____" and were asked to repeat the target word in the carrier phrase, "Now I say _____." In the second, they saw pictures of nouns and were asked to produce novel sentences containing each noun. The intelligibility of participants' pronunciations of the target words was rated by the authors. Participants in the nonword training condition showed significant gains in posttraining intelligibility, while participants in the real-word training condition did not. The authors interpreted this as indicating that when learners are able to focus on phonetic details in the absence

of lexical information (i.e., in the nonwords), they may improve their pronunciation.

Synthetic Enhancement of Training Materials. Other studies have looked at whether training effects can be amplified if relevant acoustic cues are enhanced to be made more salient. For example, Thomson (2012c) hypothesized that lengthening vowels would give listeners a better chance of noticing important differences between L2 vowel categories. Others (Iverson, Hazan, & Bannister, 2005; Jamieson & Morosan, 1986, 1989) similarly attempted to enhance cues that are thought to be helpful for hearing consonant contrasts (e.g., the acoustic third formant, or F3, is particularly important for English listeners in distinguishing /r/ from /l/). While these studies did not reveal any benefit of modified stimuli over naturally produced stimuli, phonetically "enhanced" stimuli have been associated with increased learning in other types of training studies (Escudero, Benders, & Wanrooij, 2011).

HVPT Procedures

HVPT studies have also varied in terms of their overall duration and the type of training task used. Iverson and Evans (2009), for example, provide an example of a relatively short training period, while Lambacher et al. (2005) present an example of a longer period of training. In addition, researchers have investigated the effects of the nature of the training task itself. For example, Carlet (2017) compared two training tasks: identification (e.g., Is the sound you heard A, B, or C?) and discrimination (e.g., Are A and B the same or different?). She found that while vowel perception improved after training using both tasks, participants trained with the identification task showed significantly greater improvement than those trained using the discrimination task. Hardison (2003) compared audiovisual and audio-only training, finding that audiovisual training led to significantly higher English /r/ and /l/ identification accuracy by Japanese and Korean L1 speakers.

HVPT as a Pedagogical Tool

As Bradlow (2008) notes, research on HVPT has demonstrated "conclusively that robust, linguistically-functional learning can be achieved under laboratory training conditions" (p. 299). However, the question remains whether HVPT can serve as a useful tool in the context of L2 classroom learning and teaching. The original findings of Logan et al. (1991) and Lively et al. (1993) have been reinforced by dozens of additional laboratory studies, and there is widespread agree-

ment that phonetic training with feedback using variable stimuli with regard to talker and phonetic context can contribute to improvement in the identification, discrimination, and production of L2 sounds. Researchers whose interests include the practical application of these theories to language teaching have begun to address the lag maintained between research and pedagogy. In 2004, Wang and Munro noted that "there is a significant gap between some of the key research findings of laboratory studies from the past two decades and techniques that have actually been put into practice" (p. 540). To bridge this gap, their study involved some departures from earlier HVPT studies. First, the training took place over the course of two months, a longer period than most earlier studies. Second, participants had the flexibility to decide when, how often, and how many training sessions they completed. These modifications were designed to simulate the option of extra practice outside of an English as a second language class. As with the more traditional laboratory studies, participants in the experimental condition demonstrated greater ability to identify the L2 vowel contrasts after training, while participants in the control condition did not show any improvement over two months.

Despite these promising results, very little additional research has directly investigated the application of HVPT in non-laboratory settings. A notable exception is Thomson (2011), who demonstrated that HVPT is an excellent candidate for computer-assisted pronunciation training applications. As Thomson noted, HVPT differs from regular classroom instruction and can therefore provide complementary practice. Advantages of HVPT include that it focuses attention on sounds and reduces attentional demands on meaning, it is interactive, and it involves immediate corrective feedback.

To investigate the efficacy of HVPT for improving pronunciation of English vowels for L1 speakers of Mandarin, Thomson had participants complete eight self-paced HVPT training sessions over the course of three weeks in which they learned to associate vowels with images of different types of nautical flags. Training stimuli were 10 target vowels in two CV contexts (/bV/ and /pV/) produced by 20 different native English speakers. Thomson reported that "by the end of the second training session, learners were identifying several known English vowel categories at near ceiling accuracy rates" (2011, pp. 752-753). Nineteen of the 22 participants demonstrated improvement in their pronunciation (measured using an elicited imitation task) after the training, even when imitating a new voice. Pronunciation improvement was shown in the trained CV contexts, as well as in two of the four new contexts. Thomson concluded that "when designed using a principled, research-based approach, computer-mediated train-

ing … can improve speech intelligibility without explicit practice in production" (p. 758).

In a follow-up study (Thomson, 2012c), participants from the same population participated in the same type of training and were tested to determine: (a) if their perception of the vowels improved more if stimuli were enhanced to lengthen the vowel or were selected for being maximally different from Mandarin vowel categories; (b) if perceptual improvements extended to new voices or new phonetic contexts; and (c) if perceptual improvements were retained for a month after the training. Results did not demonstrate any differences between groups, but participants improved their vowel perception after training, even on vowels produced by a new talker and on two of the four new contexts. Results from the delayed posttest showed that the improvement continued but did not increase after a month.

Thomson (2011) noted that if a web-based application were available, it "would allow endless research possibilities, as teachers and researchers could collaborate remotely, monitoring the effect of perceptual training and its impact on pronunciation, in order to improve future iterations of the software" (p. 760). To this end, Thomson has created a freely available website called English Accent Coach that enables learners to choose the English sounds they wish to work on (2012a). Training sessions using English Accent Coach are very similar to the research studies described above. Tokens are produced by 20 different voices, target sounds are provided in a wide variety of phonetic environments, and each trial contains immediate corrective feedback. In one study using this application (Thomson & Derwing, 2016), participants completed 40 training sessions at their leisure over the course of a month, with all participants demonstrating improved pronunciation after the training. Thomson (2012b) provides a useful list of ways to integrate English Accent Coach into the language classroom.

Looking Forward

The limited amount of research thus far provides evidence that the robust training effects of HVPT in laboratory studies can also occur in non-laboratory settings and have pedagogical potential for both perception and production of L2 sounds. In this section we discuss research directions to understand how to best apply HVPT in instructed language settings. Here we focus on the native and L2 sounds investigated in HVPT research, gaps in the generalizability of HVPT beyond trained stimuli, duration of training and the longevity of HVPT training effects, and the role of individual learner differences.

Need for a Wider Variety of L1-L2 Pairings

The research on HVPT was originally focused almost exclusively on Japanese speakers learning the English /r/-/l/ contrast. While research on HVPT continues to be largely English-centric, the literature covers an increasingly wide variety of languages, including: (a) native English speakers learning Mandarin tones (Perrachione et al., 2011; Wang, Jongman, & Sereno, 2003; Wang, Spence, Jongman, & Sereno, 1999); (b) native English speakers learning French vowel contrasts (Brosseau-Lapre, Rvachew, Clayards, & Dickson, 2013); (c) native Japanese and English speakers learning Hindi stop contrasts (Pruitt, Jenkins, & Strange, 2006); (d) native English speakers learning Arabic fricative contrasts (Burnham, 2013); (e) native Mandarin (Thomson, 2011) and Catalan/Spanish bilinguals (Carlet, 2017) learning English vowels; and (f) monolingual English and Spanish-English bilinguals learning Hungarian vowel contrasts (Archila-Suerte, Bunta, & Hernandez, 2016). In order to determine how HVPT works in the variety of language learning and teaching settings worldwide, more L1-L2 pairings need to be studied.

Generalizability of HVPT Training to New Phonetic Environments

Although HVPT has been shown to generalize to new talkers, it remains less clear how learning from this type of training can generalize to new phonetic environments. Lively et al. (1993) demonstrated generalization to a new phonetic environment, though Iverson et al. (2005) found that their training of the same contrast did not "fully generalize to other syllable positions" (p. 3273). Thomson's (2011) learners were able to generalize vowel production in bilabial stop-initial contexts to alveolar fricative-initial contexts but not to velar stop-initial contexts. The extent to which training in particular phonetic environments can generalize to other contexts is not known and represents a significant gap in understanding how HVPT can be used with real-world language learners.

Optimal Length of HVPT Training

Given that HVPT studies have varied in their training items and methods, many questions concerning optimal amounts of variability, or how much time it takes for learners to reach "saturation" (see, e.g., Bradlow, 2008), remain unanswered. That is, in cases in which participants have shown minimal (or no) perception or production gains, and/or an inability to generalize their learning in some way, it is unclear whether training was simply too short to result in observable improvement in participants' performance. A better understanding of

the role of training duration is crucial to effectively apply HVPT to instructed settings. Likewise, only a few studies (e.g., Bradlow et al., 1999; Lively et al., 1994; Thomson, 2012c) have demonstrated that the gains made after training are retained in the long term. Thus, the longevity of HVPT effects also requires further attention.

Individual Differences in Learners

Recent studies have documented differences among individuals in training effectiveness. A few studies have indicated that the efficacy of HVPT may depend on the perceptual sensitivity of the listener to the trained contrast (Ingvalson, Barr, & Wong, 2013; Lee, Perrachione, Dees, & Wong, 2007; Perrachione et al., 2011). Lengeris and Hazan (2010) demonstrated that the general ability to detect differences was correlated with L2 vowel discrimination and identification before and after HVPT training. They interpreted this as indicating that "some individuals are better at using spectral/acoustic information to overcome L1 biases" (p. 3767). To move forward, it is important to consider the roles that individual learner differences play in determining the impact of training. As noted by Ingvalson et al. (2013), "a better grasp on what characterizes a good nonnative speech perceiver relative to a poor one will allow for the development of better training paradigms and perhaps eliminate the variability [of results] seen in all training studies to date" (p. 6). That is, a one-size-fits-all approach to HVPT may not be appropriate, and HVPT materials and procedures ought to be tailored to aspects of the particular language-learning scenario.

Efficacy of HVPT for Real-World L2 Learners

Finally, many have suggested that HVPT is a strong candidate for applying pronunciation research in real-world L2 learning and teaching settings (Levis, 2016; Thomson, 2011) as it is low cost and can be implemented via the Internet (Thomson, 2012a). It is nonetheless necessary to empirically investigate these claims and the efficacy of HVPT for real-world learners. The stronger the empirical foundation for using HVPT in L2 instructed settings, the greater the potential benefits to language learners and teachers.

Together, laboratory studies and the application of HVPT to real-world L2 learning and teaching settings have great potential to help language learners form robust L2 sound categories with minimal extra effort on the part of the teacher—especially in the case that future research focuses on how to optimize HVPT for pronunciation learning. The improvements in perception will result also in improvements in the pronunciation of the sounds that learners find challenging.

Authors

Taylor Anne Barriuso is a PhD student in the Department of Linguistics at the University of Utah. For her dissertation, she is investigating whether the ability to learn new words containing difficult L2 sounds improves in parallel with perceptual gains from high variability phonetic training.

Rachel Hayes-Harb is a professor in the Department of Linguistics at the University of Utah. Her research concerns the role that auditory and visual input play in the acquisition of L2 phonological systems, in particular the acquisition of L2 words and their pronunciations.

References

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), *Recent research in second language phonetics/phonology: Perception and production* (pp. 2-31). Newcastle upon Tyne, England: Cambridge Scholars.

Archila-Suerte, P., Bunta, F., & Hernandez, A. E. (2016). Speech sound learning depends on individuals' ability, not just experience. *International Journal of Bilingualism*, *20*(3), 231-253.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, *167*(224), 233-277.

Bradlow, A. R. (2008). Training nonnative language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. *Phonological Second Language Acquisition*, *36*, 287-308.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English /r/and/l/: Long-term retention of learning in perception and production. *Attention, Perception, & Psychophysics*, *61*(5), 977-985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English /r/and/l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299-2310.

Brosseau-Lapre, F., Rvachew, S., Clayards, M., & Dickson, D. (2013). Stimulus variability and perceptual learning of nonnative vowel categories. *Applied Psycholinguistics*, *34*, 1-23.

Burnham, K. R. (2013). *Phonetic training for learners of Arabic* (Doctoral dissertation). Retrieved from https://repositories.lib.utexas.edu/handle/2152/21607

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America*, *54*(1973), 421-428.

Carlet, A. F. (2017). *L2 perception and production of English consonants and vowels by Catalan speakers: The effects of attention and training task in a cross-training study* (Doctoral dissertation). Retrieved from http://hdl.handle.net/10803/403758

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Attention, Perception, & Psychophysics*, *16*(3), 513-521.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*(1), 99-109.

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America*, *130*(4), EL206-EL212.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia*, *9*(3), 317-323.

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

Ingvalson, E. M., Barr, A. M., & Wong, P. C. (2013). Poorer phonetic perceivers show greater benefit in phonetic-phonological speech learning. *Journal of Speech, Language, and Hearing Research*, *56*(3), 1045-1050.

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.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267-3278.

Jamieson, D. G., & Morosan, D. E. (1986). Training nonnative speech contrasts in adults: Acquisition of the English /ð/-/θ/contrast by francophones. *Attention, Perception, & Psychophysics*, *40*(4), 205-215.

Jamieson, D. G., & Morosan, D. E. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *43*(1), 88-96.

Kuhl, P. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America, 66*(6), 1668-1679.

Kuhl, P. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6*(3), 263-285.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227-247.

Lee, J., Perrachione, T. K., Dees, T. M., & Wong, P. C. M. (2007, August). Differential effects of stimulus variability and learners' pre-existing pitch perception ability in lexical tone learning by native English speakers. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1589-1592). Saarbrücken, Germany: Universität des Saarlandes. Retrieved from http://www.icphs2007.de/

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, *128*(6), 3757-3768.

Levis, J. M. (2016). Research into practice: How research appears in pronunciation teaching materials. *Language Teaching*, *49*(3), 423-437.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358-368.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. I., & Yamada, T. (1994). Training Japanese listeners to identify English /r/and/l/. III: Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*(4), 2076-2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/and/l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874-886.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*(4), 369-390.

McClasky, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception and Psychophysics*, *34*, 323-330.

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, *18*(5), 331-340.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1496-1509.

Perkell, J. S., & Klatt, D. H. (Eds.). (1986). *Invariance and variability in speech processes.* Hillsdale, NJ: Erlbaum.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461-472.

Pisoni, D. B. (1979). On the perception of speech sounds as biologically significant signals. *Brain, Behavior and Evolution*, *16*(5-6), 330-350.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8*(2), 297-314.

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In J. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words.* Cambridge, MA: MIT Press.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America*, *119*(3), 1684-1696.

Shankweiler, D., Strange, W., & Verbrugge, R. R. (1977). Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 315-345). Hillsdale, NJ: Erlbaum.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics*, *36*(2), 131-145.

Strange, W., & Jenkins, J. (1978). Role of linguistic experience in the perception of speech. In R. D. Walk & H. L. Pick (Eds.), *Perception and experience* (pp. 125-169). New York, NY: Plenum Press.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. (1976). Consonant environment specifies vowel identity. *The Journal of the Acoustical Society of America*, *60*(1), 213-224.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, *28*(3), 744-765.

Thomson, R. I. (2012a). English Accent Coach (Version 2.3) [Web-based software]. Retrieved from https://www.englishaccencoach.com

Thomson, R. I. (2012b). English Accent Coach: Not quite a fairy godmother for pronunciation instruction, but a step in the right direction. *CONTACT English Language Learning Magazine, 38*(1), 18-24.

Thomson, R. I. (2012c). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning*, *62*(4), 1231-1258.

Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference* (pp. 88-97). Ames: Iowa State University.

Verbrugge, R. R., Strange, W., Shankweiler, D. P., & Edman, T. R. (1976). What information enables a listener to map a talker's vowel space? *The Journal of the Acoustical Society of America, 60*(1), 198-212.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033-1043.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4), 539-552.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649-3658.

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, *52*, 349-355.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Attention, Perception, & Psychophysics*, *37*(1), 35-44.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49-63.