High Variability [Pronunciation] Training (HVPT)

A proven technique about which every language teacher and learner ought to know

Ron I. Thomson
Brock University

This article is a critical research synthesis of 32 studies that used the High Variability Phonetic Training (HVPT) technique to teach learners to better perceive and produce L2 sounds. Taken together, the studies surveyed provide compelling evidence that HVPT is a very effective pronunciation training tool, and that resulting improvement is long-lasting. The analysis of this research also helps to explain why very few teachers have heard of this empirically-driven approach to pronunciation instruction: HVPT studies are largely published in technically oriented journals; few are accessible to language teachers. A variety of obstacles to the widespread use of HVPT are discussed, and some possible solutions are provided.

Keywords: L2 pronunciation, foreign accent, intelligibility, High Variability Phonetic Training (HVPT), computer assisted pronunciation training

1. Introduction

In the popular blog *Language Log*, linguist Mark Liberman (2008) lamented the common disconnect between research and practice. He described what was, for him, a twenty-year mystery: Why was High Variability Phonetic Training (HVPT), which he describes as "a simple, quick, and inexpensive technique for helping adults to learn the sounds of new languages" not more widely used in language teaching? Liberman asked some language instructors this question, expecting that they may have tried HVPT, but found no real benefit for learners. Instead, he reported that the instructors he asked had never heard of this technique to improve perception of second language (L2) vowels and consonants (segmentals).

https://doi.org/10.1075/jslp.17038.tho
Journal of Second Language Pronunciation 4:2 (2018), pp. 208–231. ISSN 2215-1931 | e-ISSN 2215-194X

In broad terms, HVPT refers to perceptual training (most often focusing on segmentals), in which the auditory training stimuli include numerous samples, produced by multiple talkers, in varied phonetic contexts. This is atypical of variation found in most pronunciation training materials. A decade after Liberman's (2008) post, HVPT remains unfamiliar to most language instructors and learners. Although at least two publically available applications of HVPT exist (Iverson & Evans, 2009; Thomson, 2018), neither application is likely to be recognized as such by end-users, who may lack awareness of the empirical rationale for these tools. In this paper, after describing the origins of HVPT, I provide a critical synthesis of thirty-two HVPT studies. I then discuss reasons for its limited use in pronunciation learning and teaching, and suggest some possible solutions.

2. Origins of HVPT

While the label HVPT seems to first appear in a study by Iverson, Hazan and Bannister (2005), as a technique, it was conceived in a series of laboratory papers aimed at answering theoretical questions about the nature of L2 speech learning (Logan, Lively, & Pisoni, 1991; Lively, Logan & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999). HVPT has also been referred to as high variability identification procedure (Lively et al., 1994), high variability perceptual training (Bradlow et al.1997) and high-variability auditory training (Iverson, Pinet & Evans, 2012), but the term HVPT appears to currently enjoy the greatest acceptance.

Although early HVPT researchers recognized the practical implications of their research, real-world application was not their primary focus. Rather, they sought to demonstrate how knowledge of L2 segmental categories evolves, to determine whether adult perceptual systems remain malleable, and under what conditions optimal learning occurs. Early HVPT research coincided with increasing debate among scholars concerning Lenneberg's (1967) Critical Period Hypothesis (e.g., Flege, 1987; Major, 1987) and its applicability to second language acquisition (SLA). Empirical evidence suggested that the relationship between age and strength of foreign accent was linear, rather than defined by an abrupt age boundary after which learning to speak without a foreign accent was impossible (e.g., Flege, Munro & MacKay, 1995). Furthermore, mounting evidence supported the notion that the quality of experience with the L2 was key to overcoming age-related effects (e.g., Flege, Bohn & Jang, 1997). Within this context, HVPT emerged as a technique that could provide focused, high quality experience, lead-

ing to measurable changes in the perceptual systems of late L2 learners, who hitherto were viewed as resistant to change, especially after the first year of exposure.

Complicating the contribution of experience, Logan et al. (1991) argued that adult knowledge of L2 sound categories varies in relation to the phonetic environments in which exemplars of particular categories occur. Given this emerging understanding of the nature of L2 phonological knowledge, they set out to identify an approach to L2 phonological training that would lead to improvement in the phonetic contexts used in training, and, importantly, extend to improvement in untrained phonetic contexts and to new talkers. In the first HVPT study, they trained six Japanese listeners to identify English /l/ from /r/, using a forced choice identification task (FCID), with feedback. Learners heard a stimulus and indicated which sound they perceived by pressing a corresponding button on a response box. Following a correct response, the system immediately proceeded to the next item. When a learner made the wrong choice, however, the system indicated that this was the case, before proceeding to the next item. Logan et al.'s stimuli included 68 minimal pairs spoken by five talkers and containing the English /l/-/r/ distinction in syllable initial, medial, or final positions. After fifteen 40-minute training sessions, the Japanese learners were significantly better at discriminating English /l/ and /r/, and this knowledge generalized to both new words and to a new talker. Learners' discrimination accuracy varied by the phonetic context in which the /l/-/r/ contrast occurred. Specifically, learners were better at identifying the contrast in syllable-final position than in syllable-initial position, before and after training. Logan et al. (1991) also found that context affected reaction time (i.e., how long it took learners to choose between /l/ and /r/).

While Logan et al.'s (1991) work provided important insights into L2 speech learning, and demonstrated that adult perceptual categories are malleable, it did not produce sufficient evidence to support the immediate large-scale adoption of HVPT as a pronunciation teaching technique. First, Logan et al. had only investigated the acquisition of a single L2 contrast, by speakers of a single L1. It was uncertain that HVPT would work equally well for other consonantal contrasts, or for vowels. Furthermore, even if the training were made available on a larger scale, it was unclear that its demonstrable effect on learners' perception of L2 sounds would transfer to production. Finally, it was not known whether the perceptual gains found in this first study would be sustained after training was concluded. Since then, many of these questions have been the focus of follow-up studies.

3. Details of studies surveyed

Including Logan et al.'s (1991) seminal research, the 32 papers chosen for analysis were identified by searching scholarly databases (e.g., Linguistics and Language Behaviour Abstracts, Scholars Portal, Google Scholar, etc.) for HVPT and its other related terms (as described in the previous section). Furthermore, some studies were identified using reference lists from recent HVPT papers, and one was added after reviewer feedback. Of the thirty-two studies selected for inclusion, 25 are from peer-reviewed journals; 6 from conference proceedings; and one is a doctoral dissertation. Of the 25 in peer-reviewed journals, 12 are published in the *Journal of the Acoustical Society of America*; 5 in psycholinguistic journals; 3 in speech communication journals; 3 in applied linguistics journals, broadly defined; and 2 in technology and computer-assisted language learning journals. Of the 6 proceedings, 4 are from phonetic sciences conferences, 1 is from a conference on speech communication, and 1 is from a conference on L2 pronunciation teaching. While some HVPT studies have undoubtedly been overlooked, these 32 provide extensive coverage of the literature.

3.1 Learner ages and language experience

Ninety-four percent of the studies report mean ages and/or age ranges of learners, but none explicitly considers age as a variable of interest. While 94 percent of the studies investigated adult learners, most participants, in most studies, were under the age of 30. A small number of older adults were sometimes included. Thirteen percent of the studies reported age ranges ending in the 30s; 25% in the 40s; 9% in the 50s; and 3% (i.e., a single study) included a 60 year-old. Only two (6%) of the studies focused on children. One investigated 16 and 17-year old high school students (Wong, 2014), while another trained 12-year old elementary school students (Hwang & Lee, 2015).

Other factors such as the age at which learners were first exposed to the target language and Length of Residence (LOR) are often reported, but as with biological age, no studies used these variables in their analyses. Studies were conducted in locations where the target L2 was spoken (e.g., Japanese speakers learning English in the USA), in foreign language learning environments (e.g., Catalan learners of English in Spain), and in experimental contexts involving naïve listeners who did not intend to learn the language targeted by the experiment (e.g., American learners of Hindi sound contrasts).

3.2 Learner L1s and target L2 sounds

The majority (84%) of the studies surveyed investigated the application of HVPT to learners of American, Canadian, and British English. Of these, 56% featured vowels exclusively, and 33% consonants exclusively. Two studies (7%) trained learners on both vowels and consonants, and another (4%) on vowels and prosody. In 61% of vowel studies, learners were simultaneously trained on a majority of possible vowel categories, while in 39% of vowel studies, learners were trained on a subset of vowels deemed most difficult. For example, some studies focused entirely on training a single vowel contrast (e.g., Ylinen et al., 2010). Consonant studies only ever focused on what researchers predetermined were difficult contrasts, rather than a larger range of sounds. For example, 80% included training on the /l/-/r/ contrast by L1 Japanese learners, and exclusively so in 60% of cases.

The five studies (16%) not focusing on English varied in terms of the learners' L1s and target languages. Two (6%) applied HVPT to the learning of Mandarin tone contrasts by American English (AE) listeners (Perrachione, Lee, Ha & Wong, 2011; Wang, Spence, Jongman, & Sereno, 1999). Another trained AE and Japanese L1 listeners to perceive a Hindi stop contrast (Pruitt, Jenkins & Strange, 2006). One trained Canadian English listeners to perceive Japanese long-short vowel contrasts (Tajima, Kato, Rothwell, Akahane-Yamada, & Munhall, 2008), and the final study trained AE listeners in the same Japanese vowel contrast as well as Japanese geminate-singleton consonant contrasts (Hirata, 2004).

3.3 Experimental groups and sample sizes

In keeping with HVPT's laboratory origins, most studies surveyed utilized comparison groups (38%), control groups (31%) or both (25%). Only two (6%) had neither comparison nor control groups. The number of learners in each experimental or control condition varied from 4 to 36 ($M=15.3$).

3.4 Number of stimulus talkers and phonetic contexts

Between 2 and 30 talkers ($M=7.2$) provided stimuli for the training procedures. While the norm is to use human talkers to produce training and testing stimuli, one study stood out by using two artificial talkers (i.e., text-to-speech) to produce stimuli (Qian, Chukharev-Hudilainen & Levis, 2018). Of the 28 studies that tested perception before and after training (the remaining four tested production only), all but one included at least one additional talker to provide test stimuli for determining if training generalized to new talkers.

3.5 Duration of training

In some cases, duration of training was specified in minutes or hours, while in others, the exact number of tokens presented to learners was also provided. In many cases, training durations were only estimated, since there were individual differences in the amount of time learners took to proceed through training tokens.

The number of training sessions and their cumulative duration varied widely from study to study. Cebrian and Carlet's (2014) learners completed four training sessions, lasting three hours in total. Nishi and Kewley-Port's (2007, 2008) participants completed nine sessions, comprising 13.5 hours. While some correspondence between training duration and the number of sounds exists, this is not always the case. For example, despite both studies providing ten hours of training, Logan et al. (1991) targeted only /l/-/r/, while Thomson (2016a) targeted ten vowels.

There is no consistency across studies with regard to length of individual sessions, nor how they are distributed over time. For instance, Tajima, et al.'s (2008) five sessions lasted 15–20 minute each over five days, while Nishi and Kewley-Port's (2007) nine sessions lasted 90 minutes, over nine days. Thomson's (2011) eight 15-minute sessions were spaced over three weeks.

3.6 Delayed post-tests

Thirty-one percent of studies conducted a delayed post-test to assess retention over periods ranging from two weeks (Flege, 1995b), to six months (Lively et al., 1994).

3.7 Testing of production

Thirty-one percent of studies assessed learners' pronunciation before and after training using recordings. While most elicited speech through read-aloud word lists, Thomson (2011) used elicited imitation via an auditory prompt in a sentence frame, while Carlet (2017) used a picture-naming task. Recordings were evaluated instrumentally, by extracting acoustic information and comparing it to native speaker values (e.g., Rato & Rauber, 2015), or by asking native speaker listeners to identify the recorded sounds they heard (e.g., Thomson, 2011).

4. Overview of findings

Of the thirty-two studies surveyed, 97% resulted in significant improvement in learners' mean perception and/or production scores for trained L2 sounds. Only Hwang and Lee's (2015) study of elementary students showed no statistically significant improvement. Their null result is not surprising. First, it is the only HVPT study surveyed in which children (12 year-olds) were trained. It is possible that HVPT training is not suitable for young learners, given its demand for focused attention. Additionally, an unusually large number of L2 contrasts were trained in Hwang and Lee's study, without a concomitant increase in the number of training sessions, relative to studies with a smaller number of contrasts. Consequently, learners may not have had enough exposure to particular L2 sounds to trigger measurable improvement. Finally, the researchers tested production only. HVPT is likely to manifest in perception first, before affecting production. Hwang and Lee (2015) observed evidence that trainees' perception had improved, but they did not measure this empirically.

4.1 Impact of individual and group differences

While age is not treated as a variable in the studies surveyed, all but two studies trained adult learners. Thus, we can safely conclude that HVPT works for adult learners. Most studies report having older learners in their age range, but only Thomson (2011) comments on an older learner's performance, indicating that a 50-year old demonstrated some of the largest gains.

Two studies investigated the impact of previous language experience on the success of HVPT, reporting that it is effective in both second and foreign language learning environments (Iverson & Evans, 2009; Iverson et al., 2012). Iverson et al. (2012) found that while those with greater target language experience were significantly better at identifying English vowels at both pre- and post-test, both groups experienced statistically equal and significant improvement across nearly all of the trained vowel categories.

Two studies examined the interaction of the learners' L1s and success with HVPT (Pruitt, et al., 2006; Iverson & Evans, 2009). In both, learners significantly improved regardless of L1 background, but L1 background affected the learners' starting points and rates of improvement. Both studies concluded that these differences stemmed from the nature of the interaction between the sounds found in the learners' L1s and the target languages.

4.2 Generalizability

Arguably, the greatest hallmark of HVPT is that perceptual learning of L2 sounds using this technique generalizes to new instances of the same sound. Every study surveyed incorporated some test of generalizability, with either new talkers, new contexts, or both.

4.2.1 *To new talkers*

Of the 28 studies that tested the effect on L2 perception before and after training (the other four tested only production), all found generalization to new talkers who produced the same words or syllables used during training. In one case, generalization occurred for all but one trained sound. Cebrian and Carlet (2014) found that training did not generalize to English /ð/ produced by a new talker. However, in their study, learners' pre-test scores in response to this talker's productions were very low relative to their performance on /ð/ training stimuli produced by other talkers. This suggests that the inability to generalize this one sound to the new talker may have been due to an unusual production or recording of that talker's test token.

4.2.2 *To the same sound targets in new words*

Tests of HVPT's generalization to new words are less common, and have met with mixed results. For example, Lively et al. (1993) reported that training with multiple talkers resulted in generalization to new words produced by new talkers, while training with a single talker failed to generalize to unfamiliar items produced by the training talker. Pruitt, et al. (2006) found that training Japanese and Americans to perceive a Hindi dental-retroflex stop contrast generalized to the same sounds in untrained contexts, as did Wong's (2014) training Chinese speakers of English /i/-/ɪ/.

The results of other studies were not as conclusive, with some showing only partial transfer of perceptual learning to the same sounds in new contexts. Iverson et al. (2005) report that improvement in the perception of English /l/ and /r/ by L1 Japanese speakers extended to new words, but primarily to words beginning with /l/s and /r/s, and not to medial position (between vowels) or within consonant clusters. Thomson (2011) found training of ten vowels in /bV/ and /pV/ contexts generalized to the pronunciation of the same vowels in /zV/ and /sV/ contexts, but not /gV/ and /kV/ contexts. In perception, the opposite pattern emerged. Vowel training in /bV/ and /pV/ contexts generalized to /gV/ and /kV/ contexts, but not to /zV/ and /sV/ contexts (Thomson, 2012a).

Others have found no transfer to new contexts. For example, Tajima et al. (2008) found that although training English speakers to perceive Japanese phone-

mic length contrasts transferred to new talkers, there was no transfer to new contexts, nor to tokens spoken at three speech rates different than that used in training. Thomson (2016a) trained learners from a variety of L1s to identify ten English vowels in /CV/ frames. Learners began with vowels presented in an /hV/ context for the first nine sessions, demonstrating significant improvement in this context. Next, 29 successive training sessions presented the same vowels in 11 different contexts (including both CV and CVC frames). Finally, training returned to the /hV/ context. Despite having received 29 sessions focussed on the same vowels in different contexts, the learners in this study showed no further improvement in their perception of vowels in the original /hV/ context. Qian et al. (2018) reported finding no significant transfer from target English vowels and consonants in trained words to the same sounds in new words. In their study, results may have been adversely affected by their unusual decision to use artificial talkers (i.e., text-to-speech) for training and generalization test stimuli.

4.2.3 *To larger stretches of discourse*

Another question is to what extent HVPT training in isolated contexts transfers to words heard in extended discourse. Hirata (2004) trained English listeners to perceive Japanese vowel and consonant length contrasts. One group were given words spoken in isolation, while another group heard the same words spoken in a carrier sentence. Hirata (2004) found that both conditions led to significant improvement. Furthermore, both groups generalized learning to the opposing context. Interestingly, she notes that the extent of transfer from training in sentence contexts to performance on isolated words was larger than the transfer of training in isolated words to sentence contexts.

4.2.4 *To untrained sound categories*

Given that sounds targeted by HVPT are never presented in isolation, but are surrounded by other non-target sounds, researchers have asked whether sounds that are not the focus of training would, nevertheless, be positively affected by training. Using identical CVC stimuli for both groups, Rato and Rauber (2015) trained one group on the vowel portion, while another group was trained on the consonant portions. The group trained on English vowel contrasts improved on all three vowel contrasts targeted, while the group that was trained on consonants also significantly improved on one of the three vowel contrasts. Such cross-transfer of HVPT to untrained sounds was also investigated by Carlet (2017). She compared two groups, one trained on vowels and one on consonants, using the same stimuli, and then examined whether the group focused on vowels learned to identify consonants and vice versa. While she did not find cross-transfer effects in most contexts, Carlet did discover that the vowel group were also better

able to discriminate voice/voiceless contrasts in coda consonants. Since discriminating whether a coda consonant is voiced or voiceless often requires attending to the preceding vowel, this is not surprising (i.e., in some sense the learners *were* attending to the vowels, despite focusing on the consonants). Ironically, the group focused on consonants did not improve in their discrimination of coda consonants.

4.2.5 *To pronunciation*

The ultimate goal of HVPT is to improve learners' pronunciation of L2 sounds. It has long been understood that normally in L2 speech development perception precedes production (Flege, 1995a; Derwing & Munro, 2015). This hints at why only a small number of HVPT studies directly measure its impact on pronunciation. Further, in keeping with the notion that perception precedes production, it is not surprising that when both are tested, significant changes in pronunciation are typically smaller than changes in perception of the same sounds (e.g., Bradlow et al., 1997; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Thomson, 2007; Iverson et al., 2012).

4.3 Retention

In addition to generalizability, another central expectation of HVPT is that learning will be retained indefinitely. As noted earlier, roughly one third of the studies surveyed included a delayed post-test, including two measuring production (Rato & Rauber, 2015; Carlet, 2017). While learning was retained at two weeks (e.g., Flege, 1995b), one month (Thomson, 2012a), two months (Rato & Rauber, 2015), three months (Bradlow et al., 1999; Wang & Munro, 2004; Nishi & Kewley-Port, 2007; Carlet, 2017), and four months (Iverson & Evans, 2009), little is known about retention beyond this period. Only two looked beyond four months. Lively et al., (1994) found that Japanese learners' improvement on English /l/-/r/ was retained at three months, but evidenced a non-significant decline at six months. Critically, while the learners' six-month scores were not significantly different from their immediate post-test scores, they were also not significantly different from pre-test scores. Wang et al. (1999) appear to stand alone in finding evidence that perceptual learning can persist, without decline, to at least six months. Since their study examined American English speakers learning of Mandarin tone contrasts, it remains uncertain whether similar retention is possible for L2 segmental categories.

4.4 HVPT versus LVPT

The results of HVPT studies are compelling, but the reason for the efficacy of the technique is not fully understood. In their second HVPT study, Lively et al. (1993) sought to confirm their hypothesis that variability in the talkers used in training leads to L2 learners' ability to generalize from training stimuli to stimuli produced by new talkers. They compared the results of a group of Japanese trained to discriminate the English /l/-/r/ contrast using six talkers against a similar group trained using a single talker – later referred to as Low Variability Phonetic Training (LVPT) by Wong (2014). Although both groups of learners demonstrated improvement in /l/-/r/ discrimination during training, only HVPT participants generalized this knowledge to a novel talker. The results of Lively et al.'s (1993) narrowly focused study were not replicated until Perrachione, et al.'s (2011) study of AE learners of Mandarin tone. As with Lively et al. (1993), only the HVPT condition resulted in generalization to new talkers and words. Wong (2014) later replicated this finding with Cantonese learners of English /e/-/æ/, who generalized to new talkers and words in only the HVPT condition. While LVPT may also be conceived of as limiting variability in terms of the number of phonetic environments in which target sounds are presented, none of the studies surveyed here explicitly controlled for such variability.

4.5 HVPT versus other techniques

Similarly, teachers and learners should be assured that HVPT is at least as good, if not better, than competing techniques. Iverson et al. (2005) compared HVPT and three other approaches for teaching Japanese L1 speakers the English /l/-/r/ contrast. These included, (1) primary cue enhancement, (2) perceptual fading, and (3) secondary cue variability. Primary cue enhancement is a technique where software is used to manipulate and enhance primary phonetic cues associated with the correct identification of a particular speech sound. In the case of /l/ and /r/, the important primary cue distinguishing this pair is the third formant (a region of noise in the acoustic signal that differs for /l/ and /r/). This third formant is typically ignored by speakers of Japanese, because their language does not have a similar phonological contrast, but it becomes more salient through cue enhancement. The perceptual fading condition was identical to primary cue enhancement at the beginning of training, in that the third formant of /l/ and /r/ tokens were exaggerated to a more salient level during initial training sessions. Once listeners discriminated the contrast, subsequent training used the same tokens, but with the third formant faded back to natural speech levels. Iverson et al.'s secondary cue variability approach involved beginning with reliable secondary acoustic cues associated

with /l/-/r/ discrimination (including the second formant and segment duration) before increasing secondary cue variability to natural levels later. Iverson et al. (2005) found all three approaches to training led to equal and statistically significant improvement. While this study may seem to raise questions about HVPT's superiority, the researchers concluded that given the labor-intensive nature of the other techniques, which require extensive preparation time for stimuli, HVPT is preferable.

4.6 Limits on variability and its effectiveness

While there is broad agreement that variability in training stimuli leads to generalizable learning, evidence suggests that unconstrained variability may be counterproductive. Perrachione et al. (2011) found that learners' success with HVPT depends on their perceptual aptitude. In their study, low perceptual aptitude learners had difficulty learning Chinese tonal contrasts when talker variability was incorporated into every training session. However, when they blocked each training session by a single talker, introducing new talkers session-by-session, low perceptual aptitude learners performed as well as high perceptual aptitude learners. Critically, high perceptual aptitude learners performed equally well whether talkers varied within sessions or only across sessions. As such, Perrachione et al. (2011) concluded that blocking by talker is the best approach.

It remains unknown whether the within trial variability that led low perceptual aptitude learners to struggle with Mandarin tone learning would similarly affect HVPT as it applies to segmental learning. Blocking training by talkers is a common practice in other HVPT studies, but this only confirms that low within-trial talker variability is effective, including for learners who are known to have low perceptual acuity (see Lengeris & Hazan, 2010). It does not prove that within-trial talker variability is detrimental, nor that it might not be advantageous. Thomson's (2012a) study of English vowel learning used 20 talkers per trial, yet almost all learners significantly improved. It can reasonably be assumed that some low perceptual aptitude learners were in that sample.

Another concern about the nature of variability pertains to the optimal number of sounds to be targeted within a single training session. For vowels, Nishi and Kewley-Port (2007) provide compelling evidence that training learners in the perception of larger portions of the vowel space is more effective than training them on only a predetermined subset of contrasts. They found that training L1 Japanese learners to perceive nine American English vowels was more efficient than training them to identify a subset of three difficult vowels; with the same amount of training, the group given nine vowels improved across all vowels, while the group trained on three vowels only improved on those three, and to no greater

extent than the group trained on nine. In a follow-up study, Nishi and Kewley-Port (2008) trained L1 Korean learners on the same nine AE vowels, but compared a group that began training on three difficult vowels, before being introduced to the remaining six, with two groups that began training with all nine vowels. The groups that began with nine vowels significantly improved across all categories, while the group that began training on the difficult three-vowel subset did not. Despite this evidence, other researchers continue to use approaches that include subset training to significant effect (e.g., Iverson & Evans, 2009; Shin & Iverson, 2013).

4.7 Modifications to canonical HVPT

As noted earlier, in its original and most basic form, HVPT uses natural training stimuli produced by multiple talkers, in multiple phonetic contexts. Further, it typically uses a forced choice identification (FCID) paradigm, in which listeners hear a stimulus and indicate, from among two or more alternatives, what they believe they heard. Several studies have further investigated HVPT by manipulating either the presentation or quality of the stimuli.

Flege (1995b) examined whether same or different discrimination tasks (AX tasks) are more effective than FCID tasks. Flege found that for training word final English /d/-/t/ contrasts, AX tasks and FCID tasks worked equally well. In contrast, Carlet (2017) investigated the same question for the training of English vowels and found that while both approaches resulted in improvement, the FCID led to significantly larger gains than the AX task. Taken together, these results suggest FCID should remain the method of choice.

Hazan, Sennema, Iba, and Faulkner (2005) explored whether adding a visual modality to HVPT training, in which learners not only heard the target sounds, but also saw the speaker producing them, would lead to greater gains than auditory training alone. They found the addition of visual cues for the English /v/-/b/ contrast led to significantly greater improvement than audio alone. In contrast, no significant benefit of visual information was found for training the /l/-/r/ contrast. They suggested that this mixed result is due to the more visually distinct nature of the labial sounds.

Thomson (2011) sought to improve on the success of HVPT by modifying the auditory stimuli themselves. One group of learners was trained to perceive English vowels using naturally produced CV tokens, while another group was trained on synthetically manipulated tokens, in which the vowel portions were doubled in length to provide learners with exaggerated cues to vowel identity. While both groups improved, manipulating the vowels' durations did not lead to significantly greater improvement between groups. While Thomson (2007)

found some evidence that lengthened vowel stimuli are easier to identify perceptually, lengthened vowel training did not affect learners' responses to natural, non-lengthened test items.

Given the phonetic orientation of perceptual training, Thomson and Derwing (2016) investigated whether training on nonsense words would lead to greater improvement than training with real words, arguing that nonsense words are more likely to force learners to process sounds phonetically. They found that participants trained to perceive ten English vowels in CV syllables significantly improved in their pronunciation of 70 real words containing the same vowels. Another group, trained to perceive the same vowels, but using 70 real words, did not significantly improve in their pronunciation of those same words. This finding is consistent with Nishi & Kewley-Port (2007) and Carlet (2017), who while not investigating the same question, used nonsense words to train vowel perception. Both found that training generalized to vowels in both nonsense words and real words.

5. Critique of methods and gaps in findings

While this survey of 32 studies indisputably demonstrates the effectiveness of HVPT, the lack of any identifiable best approach to HVPT is self-evident, and may partially explain why it has not gained broader acceptance. In this section, I underscore some methodological shortcomings of HVPT studies, and highlight remaining gaps in our understanding of this promising technique.

5.1 Individual differences

It was noted earlier that despite providing many details about learners' backgrounds, researchers rarely interpret results with reference to that information. Some notable exceptions include Nishi and Kewley-Port (2007, 2008) and Thomson (2011, 2012a), who provide learning gains for individual learners, in addition to group means. In doing so, they demonstrate that nearly every learner improved. They do not, however, offer any potential explanations for why a few individual learners did not improve. Other studies highlighting individual differences include Lengeris and Hazan (2010) and Perrachione et al. (2011), who examined the perceptual aptitude of learners in relation to their success with HVPT. Finally, Iverson et al. (2012) examined differences between learners in foreign language learning contexts relative to those where the L2 was the primary language of the environment, demonstrating that HVPT works in both contexts.

Some major gaps in our understanding of how individual differences influence the success of HVPT include the effects of age, motivation, and learning styles. For example, is there a limit on HVPT's efficacy with much older learners? What role does age of first exposure play? Would gamified HVPT paradigms be more effective with young learners? Can learners be intrinsically motivated to use HVPT, or does it require extrinsic motivation to succeed (many studies have offered participants money or course credit)?. Does HVPT work better for some learning styles than others? Does ambiguity tolerance play a role?

5.2 Learners' L1s and target L2 sounds

HVPT studies are predominantly focused on English learners, and specifically, standard varieties of English. By necessity, any given application of HVPT must be constrained to a single variety, since categorical boundaries being learned are, by definition, dialect-specific. However, there are likely opportunities to test HVPT in contexts where other varieties are spoken. Clearly, more research is needed to determine if HVPT is generalizable across a wider range of L2s, by speakers of a wider range of L1s.

HVPT studies have targeted training of several L2 vowels. For L2 consonants, however, there is a much larger gap in our understanding of its effectiveness. A majority of consonant studies have focused on a single contrast, English /r/-/l/. To make matters worse, these two sounds have vowel-like characteristics, making them poor candidates from which to form conclusions about consonants in general.

Another methodological concern is that, in many studies, decisions about which sounds to train are based on a priori assumptions regarding which L2 sounds are problematic for a specific group of learners. These assumptions often rely on a contrastive analysis approach comparing the L2 with the learners' L1, or are even based on intuition. One case where this might lead to erroneous conclusions is English /l/-/r/. There is a long tradition of assuming that Japanese speakers confuse these two sounds, yet, in a controlled experiment, Guion, Flege, Akahane-Yamada, and Pruitt (2000) discovered that some English /r/ tokens are also confusable with English /w/, suggesting that HVPT training could include a three-way contrast. Avery and Ehrlich's (1992) widely used text for pronunciation teachers predicts that Mandarin speakers will confuse English /i/ and /ɪ/. Yet, Thomson (2007) found that these two sounds were almost never confused by Mandarin speakers, although /ɪ/ was often confused with /ɛ/. Finally, even learners of the same L1 background can differ in terms of what L2 sounds they find difficult (Derwing & Munro, 2015). A solution to addressing learner differences is to create open-ended HVPT systems allowing individual learners to test their con-

fusions across multiple categories, to determine where they need improvement. Thomson (2018) English Accent Coach includes almost all English sounds, allowing individual learners to quickly identify which ones present the greatest challenge, and to customize their training accordingly.

5.3 Experimental groups and sample sizes

While most studies utilized either a comparison or a control group, 6% did not. As Thomson and Derwing (2015) emphasize, an ideal study should include a control group to ensure that gains in learning are the result of training. At the very least, researchers should provide a rationale for why learners would be unable to improve to the same extent without HVPT.

Sample sizes in HVPT studies are sometimes smaller than they should ideally be. This is likely a consequence of the labor-intensive nature of these studies, and the use of convenience sampling. As HVPT starts using the cloud more frequently, running experiments on the internet will allow access to many more participants (e.g., Thomson's (2018) English Accent Coach).

5.4 Number of stimulus talkers and phonetic contexts

At one extreme, Nishi and Kewley-Port (2007, 2008) used only two talkers in training, while at the other extreme, Thomson and Derwing (2016) used 30. The optimal number of talkers for maximum effectiveness is still unknown. Future studies should explore this question in detail as the answer will determine how difficult it will be to develop HVPT applications to provide the greatest benefit to learners. Similarly, little is known about how many phonetic contexts should be utilized for greatest effect, and whether this depends on the specific target sound, or sounds, being learned.

5.5 Presentation of stimuli

Numerous studies have followed Logan et al.'s (1991) and Lively et al.'s (1993) lead and blocked training sessions by talker, instead of using multiple talkers within a single session. Only one study surveyed investigated whether using multiple talkers within the same session is advantageous to learning. While Perrachione et al. (2011) found evidence that blocking sessions by individual talkers is the better strategy, they investigated the learning of synthetic Mandarin tone by AE learners. Thus, it would be unwise to assume that their findings apply to learning L2 vowels and/or consonants. In fact, there is ample evidence to suggest that learners are successful at learning vowels, even when there are as many 30 talkers per trial

(e.g., Thomson, 2016a; Thomson & Derwing, 2016). Conversely, it is possible that the learners in such studies would have benefited even more had the training been blocked by talker.

There is strong evidence to support the belief that using the Forced Choice ID (FCID) paradigm to present HVPT stimuli is more effective than using an AX discrimination approach. This appears to be the case for consonants (Flege, 1995b) and vowels (Rato & Rauber, 2015; Carlet, 2017). As such, unless the purpose is to replicate previous findings, HVPT ought to use FCID.

Finally, questions remain about the use of adaptive training, which refers to changes in the structure of training in response to individual learner performance. While several studies utilized some form of adaptive training (e.g., Pruit et al., 2006; Iverson & Evans, 2009; Lengeris & Hazan, 2010), none compared it to non-adaptive training. Future research should determine how an adaptive learning approach can best be applied in the context of HVPT.

5.6 Modifications to canonical HVPT

Unless a study explicitly tests the benefit of modifying a basic approach to HVPT, new variables should not be introduced. For example, Thomson's (2011, 2012a) use of nautical flags as response choices, instead of phonetic or orthographic symbols, introduced a variable that makes it difficult to compare his studies with others. While learners in these studies showed large gains, it is unclear whether the decision to use nautical flags affected learning. Wang and Munro (2004) used a mix of traditional HVPT (four talkers), and two types of synthetic speech in their training program, while Cebrian and Carlet (2014) incorporated a variety of tasks, including an awareness raising task, discrimination training and FCID. While using multiple types of stimuli and/or tasks may have helped ensure improvement, there is no way of knowing which particular stimuli or tasks, and in what combination, were ultimately responsible for gains.

5.7 Duration of training

The duration of training necessary to reach maximum improvement also remains unknown. None of the studies surveyed indicated that learners hit a ceiling in their improvement. If there is a ceiling, is it short of native speaker norms? Further, is it reached abruptly? If not, are the largest gains made early in training, followed by a decline in the rate of improvement?

More must also be learned about the optimal duration of individual sessions. When does learner fatigue set in? Do longer sessions result in a declining return on time spent, relative to shorter sessions spread over more days or weeks?

5.8 Delayed post-tests

Only 30% of the studies surveyed included a delayed post-test. Carefully devised longitudinal studies are necessary to determine whether learning is permanent, and to what extent it is manifested in production later on in learners' L2 development. Maintaining gains from HVPT may depend on the extent to which learners continue to study the language after training is complete. It is also possible that a certain threshold in ability must be reached before changes become permanent.

5.9 Testing of production

While it is encouraging that more HVPT studies measure both perception and production, care should be taken in how pronunciation is assessed. First, speech elicitation tasks should reflect a variety of speaking contexts, and ideally include spontaneous speech. As with the broader field of pronunciation instruction (see Thomson & Derwing, 2015), HVPT studies over-rely on reading tasks to assess learners' pronunciation. More studies should utilize elicited imitation (see Thomson, 2011) or picture naming tasks (see Carlet, 2017) for assessment that realistically reflects learners' ability to pronounce target L2 sounds.

In addition to greater care in the selection of speech tasks, recordings should be assessed by human judges. While comparing acoustic measures of L2 productions with those of native speakers provides an indication of change (e.g., Rato & Rauber, 2015), only human raters can accurately determine to what extent that change is on target, and whether it contributes to the speakers' intelligibility and comprehensibility (processibility) for listeners, which ought to be the ultimate goal (Thomson & Derwing, 2015).

6. Discussion

This article began with Liberman's (2008) questions about why High Variability Phonetic Training (HVPT), which he describes as "a simple, quick, and inexpensive technique for helping adults to learn the sounds of new languages" is not more widely used in language teaching, and why most teachers have never heard of it. This review of 32 HVPT studies, half of which were published since Liberman's (2008) blog, provides numerous clues.

First, the term HVPT may be a deterrent to language teachers, who often lack training in phonetics (Thomson, 2013). Implicit in the title of this article, it may be helpful to maintain the same initialism, while changing its full form

to High Variability *Pronunciation* Training. This would be a more user-friendly term, reflecting its desired application in the real world.

Second, most studies describing HVPT's effectiveness are relatively inaccessible to teachers and learners. Over half the studies surveyed are published in highly technical phonetics journals, while most others are in niche venues, or conference proceedings that are unlikely to be accessed by pronunciation instructors. While a few are written for applied linguists, only one is published in a venue directed at pronunciation teachers and researchers. Remarkably, despite being a computer-mediated technique, only one is published in a journal devoted to computer-assisted language learning. HVPT researchers should relay the technical and scientific details of HVPT in teacher-friendly descriptions, if this gap is to be bridged (e.g., Thomson, 2012b, 2016b).

Third, a lack of clear agreement on what constitutes best practice in HVPT remains elusive. The many variations of HVPT found here makes choosing a specific approach overwhelming. While we await more definitive answers to some of the gaps in our understanding, HVPT systems can be built which allow users to manipulate some yet-to-be-proven parameters. For example, individual users should be able to control which sounds are the focus of training; whether to block training sessions by single talkers or multiple talkers; whether to present stimuli in isolated words or in sentence contexts; whether to allow repetition of stimuli before they must respond; whether to use IPA characters as response labels, or something else; how long each session should last, etc. Other aspects of HVPT training are largely established as best practice. For example, HVPT systems should include multiple talkers, producing target sounds in multiple contexts/words. In addition, the training should use an FCID paradigm, and provide learners with feedback on the accuracy of their responses, although variation in the nature of feedback given has not been the focus of much research.

Fourth, Liberman's (2008) claim that HVPT is a simple, quick and inexpensive technique is inaccurate. He may have been responding to the simplicity of a very narrow HVPT study of two English sounds. As a complete system, comprising many L2 sounds, for learners of varied L1 backgrounds and L2 experience, HVPT quickly becomes much more complex. This soon adds up to thousands of tokens, which must be recorded and screened for accuracy. Next comes the issue of how to implement the training. Until Rato, Rauber, Kluge, and dos Santos' (2015) developed TP (Tests of Perception) a free Windows®-based program for creating perceptual training tasks, there were no suitable and free programs to help teachers create such training. While TP is relatively easy to use, it takes effort to develop individual perceptual tasks, and requires access to computers. Other applications, such as Thomson's (2018) English Accent Coach, and Iverson and Evans (2009) Vowel Trainer, provide ready-to-use HVPT options, but they

are limited to the stimuli with which they are prepackaged. Further, Thomson's (2018) application is web-based, requiring a reliable and high speed internet connection, while Iverson and Evan's (2009) application must be installed locally on a Windows® computer.

Fifth, even these publically available HVPT applications garner limited use relative to the demand for pronunciation training. While this is due, in part, to a lack of teacher and learner familiarity with the science behind them, a lack of advertising funds for promotional materials is also a root problem. Simply put, most teachers and learners do not know these applications exist. They are in direct competition with often-times unscrupulous and predatory purveyors of so called 'accent reduction' or 'accent modification' services (Thomson, 2014; Derwing & Munro, 2015). While many such services are not effective, their for-profit nature provides advertising revenue allowing them to be ranked higher in internet search engines.

Sixth, mainstream language teachers may be reluctant to introduce HVPT to learners, since it does not easily lend itself to incorporation in traditional classrooms. Being computer-based, it requires language lab time. Where lab facilities are unavailable, or lab time is limited, teachers need to assign HVPT as homework. HVPT apps for mobile devices would allow for this, since most students have smartphones, but monitoring whether such homework is done properly is difficult. Teachers must also find ways of incorporating HVPT into the larger curriculum, and extending it to communicative activities.

Seventh, independent use of HVPT technologies outside of the classroom requires that students be highly motivated. Current approaches to HVPT tend to be boring, and thus demotivating. One exciting avenue for future development is through collaboration with the gaming industry. Lim and Holt (2011) demonstrated, on a small scale, that perceptual learning can occur incidentally during video game play. For example, they created a game in which users had to navigate a 3D tunnel while encountering four alien types, each of which had a unique sound. Integration of HVPT stimuli into such games may be appealing to some learners.

7. Conclusion

While HVPT's effectiveness for improving learners' perception and pronunciation of L2 sounds is well documented, many obstacles remain for more extensive use in language teaching. As further HVPT research addresses some of the outstanding questions identified here, a coherent set of best practices for HVPT can be established. Even so, more work is needed to bridge the gap between research and prac-

tice. This requires better cooperation between researchers and teachers, and also between researchers and programmers. It also requires a willingness on the part of researchers to write for a teacher audience. On the bright side, given the rapid move toward cloud-based applications, obstacles associated with the accessibility of large-scale HVPT applications are likely to disappear soon.

Acknowledgements

I would like to acknowledge the organizers of the 2016 Pronunciation in Second Language Learning and Teaching Conference, for inviting me to the Roundtable on the role of technology in future pronunciation research and instruction, which was sponsored by *Language Learning*. This research synthesis is an outcome of my presentation there. I would also like to thank Tracey Derwing and two anonymous reviewers, for helpful feedback on earlier drafts of this article.

References

Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation*. Oxford: Oxford University Press.

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.

Carlet, A. (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 (Unpublished doctoral dissertation). Universitat Autònoma de Barcelona.

Cebrian, J., & Carlet, A. (2014). Second-language learners' identification of target-language phonemes: A short-term phonetic training study. *Canadian Modern Language Review*, 70(4), 474–499.

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

Flege, J. E. (1987). A critical period for learning to pronounce foreign languages? *Applied Linguistics*, 8(2), 162–177.

Flege, J. (1995a). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E. (1995b). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, 16, 425–442.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, 97(5), 3125–3134.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437–470.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, 107(5), 2711–2724.

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.

Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *The Journal of the Acoustical Society of America*, 116(4), 2384–2394.

Hwang, H., & Lee, H. Y. (2015). The effect of high variability phonetic training on the production of English vowels and consonants. In the Scottish Consortium for ICPhS (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Paper number 466. Glasgow, UK: Glasgow University.

Iverson, P., & Evans, G. B. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *Journal of Acoustical Society of America*, 126(8), 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.

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(1), 145–160.

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.

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.

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, 2(12), 59–67.

Liberman, M. (2008, July 6). HVPT. [Blog post]. Retrieved from <http://languagelog.ldc.upenn.edu/nll/?p=328>

Lim, S. J., & Holt, L. L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science*, 35(7), 1390–1405.

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.

Major, R. C. (1987). Foreign accent: Recent research and theory. *International Review of Applied Linguistics in Language Teaching*, 25(4), 185–202.

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.

Nishi, K., & Kewley-Port, D. (2008). Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research*, 51(6), 1480–1493.

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.

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.

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

Rato, A., Rauber, A. S., Kluge, D. C., & dos Santos, G. R. (2015). Designing speech perception tasks with TP. *Investigating English pronunciation: Trends and directions.* In J. A. Mompean & J. Fouz-González, (Eds.), *Investigating English pronunciation: Trends and directions.* (pp. 295–313). Houndmills: Palgrave Macmillan.

Rato, A., & Rauber, A. (2015). The effects of perceptual training on the production of English vowel contrasts by Portuguese learners. In the Scottish Consortium for ICPhS 2015. (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.* Paper number 656. Glasgow, UK: Glasgow University.

Shin, D. J., & Iverson, P. (2013). Training Korean second language speakers on English vowels and prosody. In *Proceedings of Meetings on Acoustics ICA* 2013, 19(1). Melville, NY: Acoustical Society of America.

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., & Munhall, K. G. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *The Journal of the Acoustical Society of America*, 123(1), 397–413.

Thomson, R. I. (2007). Modeling L1/L2 interactions in the perception and production of English vowels by Mandarin L1 speakers: A training study (Unpublished doctoral dissertation). University of Alberta.

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

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

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

Thomson, R. I. (2013). ESL teachers' beliefs and practices in pronunciation teaching: Confidently right or confidently wrong? In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference.* Aug. 2012. (pp. 224–233). Ames, IA: Iowa State University.

Thomson, R. I. (2014). Myth 6: Accent reduction and pronunciation instruction are the same thing. In L. Grant (Ed.). *Pronunciation myths: Applying second language research to classroom teaching.* (pp.160–187). Ann Arbor, MI: University of Michigan Press.

Thomson, R. I. (2016a). Does training to perceive L2 English vowels in one phonetic context transfer to other phonetic contexts? *Canadian Acoustics*, 44(3), 198–199.

Thomson, R.I. (2016b). Learning to perceive English sounds using computer-assisted pronunciation training. *AL Forum, November*, 2016. Alexandria, VA: TESOL International Association.

Thomson, R.I. (2018). English Accent Coach [Computer program]. Version 2.3. Retrieved from <www.englishaccentcoach.com>

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

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*. Oct. 2015. (pp. 88–97). Ames, IA: Iowa State University.

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.

Wong, J.W.S. (2014). The effects of high and low variability phonetic training on the perception and production of English vowels/e/-/æ/by Cantonese ESL learners with high and low L2 proficiency levels. In *the Fifteenth Annual Conference of the International Speech Communication Association*. Singapore, 14–18 September.

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2010). Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *Journal of Cognitive Neuroscience*, 22(6), 1319–1332.

Address for correspondence

Ron I. Thomson
Department of Applied Linguistics
Brock University
St. Catharines, ON
Canada
rthomson@brocku.ca