THE L2 ACQUISITION OF PHONEMES AND ALLOPHONES
UNDER VARIOUS EXPOSURE CONDITIONS

by

Taylor Anne Barriuso

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics
The University of Utah
August 2018
The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of ________________________________ Taylor Anne Barriuso

has been approved by the following supervisory committee members:

Rachel Hayes-Harb, Chair 5/4/2018 Date Approved

Shannon Barrios, Member 5/4/2018 Date Approved

Johanna Watzinger-Tharp, Member 5/23/2018 Date Approved

Ron Thomson, Member 5/26/2018 Date Approved

John Levis, Member 5/26/2018 Date Approved

and by ________________________________ Scott Jarvis, Chair of

the Department of ________________________________ Linguistics

and by David B. Kieda, Dean of The Graduate School.
ABSTRACT

One of the biggest challenges facing adult language learners is acquiring the sound system of the second language. While it has been shown that adults can acquire novel phones in their second language, the relevant features of the input available for use remain to be determined. Two proposed features of the input are investigated in this dissertation.

The first study investigates the role of phonological distributional information in the acquisition of allophones. Native English speakers were exposed to the Spanish [b]-[β] alternation in either overlapping or complementary distribution. It was expected that if participants inferred the phonological relationship from the phonological distribution, participants exposed to the pair in overlapping distribution would outperform participants exposed to the pair in complementary distribution on an ABX task. The two groups perform the same on the discrimination task, suggesting that phonological distributional information alone may be insufficient for adult allophone acquisition.

The second study, investigating whether perceptual gains from high variability phonetic training (HVPT) can generalize to higher level tasks, employs a freely available online implementation of HVPT (English Accent Coach [EAC]) with actual second language English (ESL) learners in two scenarios: part of ESL class sessions or at the times and locations of the learners’ choosing (simulating extracurricular practice). In each setting, ESL students trained using EAC and a control group of ESL students complete a
pretest and a posttest, each containing a discrimination task and an artificial-lexicon learning task. In the classroom setting, students from both groups demonstrated no improvement on either task from pretest to posttest. In the extracurricular setting, the HVPT group outperformed the control group on the discrimination posttest, but there was no difference for either group on the lexical posttest. Together, these results suggest that (1) efficacy of HVPT for perception improvement relies on adherence to some protocols and (2) the ability to use perception knowledge for a lexical task may not develop in parallel with perception.

The results are discussed with regard to their implications for language teaching and their contributions to our understanding of what features of the input adults are able to take advantage of for L2 phonological development.
To Mr. Reinsmoen,

who told wee 4th grade me
someday he would have a book on his shelf
with my name on it.
TABLE OF CONTENTS

ABSTRACT .......................................................................................................................... iii

LIST OF TABLES ................................................................................................................ viii

LIST OF FIGURES ............................................................................................................ ix

Chapters

1. HIGH VARIABILITY TRAINING AS A BRIDGE FROM RESEARCH TO PRACTICE................................................................. 1
   Abstract .......................................................................................................................... 2
   Introduction .................................................................................................................... 2
   A Brief History of HVPT ............................................................................................... 3
   HVPT and Second Language Acquisition ..................................................................... 7
   HVPT as a Pedagogical Tool ....................................................................................... 14
   Looking Forward ........................................................................................................ 17
   References ..................................................................................................................... 21

2. THE ROLE OF PHONOLOGICAL DISTRIBUTIONAL INFORMATION ON THE ACQUISITION OF L2 ALLOPHONES ......................... 27
   Abstract .......................................................................................................................... 27
   Introduction .................................................................................................................... 28
   Background ................................................................................................................... 29
   The Present Study ....................................................................................................... 33
   Discussion ...................................................................................................................... 39
   References ..................................................................................................................... 42

3. HIGH VARIABILITY PHONETIC TRAINING AND WORD LEARNING: AN INVESTIGATION OF THE GENERALIZABILITY OF PERCEPTION GAINS TO HIGHER-LEVEL TASKS FOR SECOND LANGUAGE LEARNERS ................. 45
   Introduction ................................................................................................................... 45
   Background .................................................................................................................... 46
   The Present Studies ...................................................................................................... 52
   Experiment 1 ................................................................................................................. 52
   Experiment 2 ................................................................................................................. 69
General Discussion ........................................................................................................... 79
References ....................................................................................................................... 85

4. THE STATE OF THE FIELD AND FUTURE DIRECTIONS ......................... 90
   Introduction .................................................................................................................. 90
   Distribution-Based Learning and L2 Allophonic Acquisition ................................. 90
   High Variability Phonetic Training and L2 Phonological Acquisition ................. 96
   References ................................................................................................................. 105
LIST OF TABLES

Tables

2.1 Stimuli heard in each exposure group................................................................. 36

3.1 Self-reported demographic data from the 28 coaches who elected to participate ..... 56

3.2 Word learning stimulus items.................................................................................. 60

3.3 AXB stimulus items................................................................................................. 62

3.4 Coaches who completed each task.......................................................................... 63

3.5 Self-reported demographic data from the 20 participants with usable data........... 70

3.6 Summary of results from ANOVA of group x task type x time.............................. 71

3.7 E2 AXB mean d-prime scores by time and token type.......................................... 74

3.8 E2 AXB mean d-prime scores by contrast and time............................................. 75

3.9 Mean AXB d-prime scores for all groups by token type and time.......................... 81
LIST OF FIGURES

Figures

2.1 Boxplot of mean proportion correct by exposure group and contrast ..................... 37
2.2 Boxplot of d-prime scores by exposure group and contrast.................................. 38
3.1 An HVPT training session on the website englishaccentcoach.com....................... 58
3.2 The three phases of the word learning task.......................................................... 61
3.3 E1 AXB mean d-prime scores by group and time.................................................. 65
3.4 Scatterplot of the number of tokens that each E1 HVPT participant was trained on and their improvement on the AXB task with a superimposed line of fit......................... 65
3.5 E1 word learning mean d-prime scores by group and time.................................... 67
3.6 E2 AXB mean d-prime scores by group and time.................................................. 73
3.7 E2 AXB mean d-prime scores by contrast and time............................................. 75
3.8 E2 WL mean d-prime scores by group and time.................................................... 77
3.9 Plot of the five target vowels as produced by the three testing voices (solid line) and three of the 20 EAC voices (dashed line) ................................................................. 83
The purpose of this dissertation is to address the broad question of how specific features of L2 input can be used by adult second language learners for phonological development. Chapter 1 is a review of high variability phonetic training (HVPT) as a field of inquiry with potential to bridge some of the gap between research and practice in L2 phonological acquisition. Chapter 2 will look at the role of phonological distributional information in the acquisition of L2 allophonic alternants. Chapter 3 will investigate whether HVPT can lead to greater accuracy not only on perceptual discrimination tasks, but also on higher level word learning tasks by actual ESL learners. Finally, Chapter 4 will provide a discussion of the contributions of this dissertation to the field and future directions.

1 The manuscript presented in this chapter is an adapted version of a manuscript published in *The CATESOL Journal Special Issue on Pronunciation*, guest edited by John Levis, Donna Brinton, and Ana Wu. The manuscript presented here was written for an audience of scholars of L2 speech acquisition and processing, while the manuscript published in *The CATESOL Journal* was written for an audience consisting primarily of language teachers.
Abstract

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) phonological acquisition. 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 what is learned from HVPT to new situations, in addition to methodological considerations that may promote optimal learning. In primary focus is the relatively limited literature 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.

Introduction

Learning the pronunciation of a new language means learning both to produce and perceive new speech sounds. 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. Decades of speech research, however, have provided evidence of the efficacy of a technique called high variability phonetic training (HVPT) for increasing listeners’ sensitivity to nonnative sounds. The hallmark of HVPT is that it uses multiple talkers to provide the variability inherent in speech produced by different people, which is presumed to help learners sort out which dimensions of variability are and are not
important for differentiating the novel L2 sounds.

In this article we review selected HVPT research for the purpose of understanding 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. Levis 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 empirical investigation, language teachers and curriculum designers may benefit from taking advantage of the knowledge established through this robust line of laboratory 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. Following Liberman, Harris, Hoffman, and Griffith’s (1957) finding that (native) speech sounds are perceived categorically, many new questions arose including: (a) whether this ability is innate or learned (e.g., Kuhl, 1979), (b) if it is language-specific or perception-general (e.g., Eimas, 1974), (c) what factors contribute to category development (e.g., Eimas & Corbit, 1973), (d) how flexible these sound-to-category mappings are once they have been formed (e.g., Verbrugge, Strange, Shankweiler, & Edman, 1976), and (e) the role of phonetic environment on segment
identification (e.g., Strange, Verbrugge, Shankweiler, & Edman, 1976).

Much of this work focused on what was then referred to as the *constancy problem* (Kuhl, 1979, 1983; Shankweiler, Strange, & Verbrugge, 1977); that is, listeners’ highly accurate perception of (native) speech despite the inherently variable nature of the speech signal. Over time, speech perception researchers became interested in the consequences of speech variability not only for native speech processing by adults (see, e.g., Pisoni, 1979 for a review), but also for acquisition and processing by infants and “cross-language” speech processing by adults (see e.g., Perkell & Klatt, 1986).

Research on cross-language speech perception revealed that listeners do not perceive nonnative contrasts in the same way as native listeners (e.g., Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973) or with the same accuracy as native listeners (e.g., Miyawaki et al., 1975; Werker & Logan, 1985). A body of research developed to probe the phenomenon of cross-language speech perception, focusing on questions concerning the time-course of the emergence of language-specific speech perception in infancy (e.g., Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984; see also Best, 1994 for a review) and the difficulty of perceiving nonnative phonological contrasts (e.g., Goto, 1971; MacKain, Best, & Strange, 1981; Werker & Tees, 1984). In the 1980s some researchers began asking whether listeners could be trained to perceive nonnative speech sounds (e.g., Jamieson & Morosan, 1986; McClasky, Pisoni, & Carrell, 1983; Pisoni, Aslin, Perey, & Hennessy, 1982; Strange & Dittman, 1984; Strange & Jenkins, 1978). By the mid-1990s, a critical aspect of the difficulty associated with nonnative speech perception had been identified: determining which dimensions of acoustic variability are relevant for distinguishing difficult contrasts (Pisoni, Lively, & Logan, 1994).
In a seminal study, Logan, Lively, and Pisoni (1991) hypothesized that training using stimuli that involved variability in the forms of different talkers and different phonetic environments might help learners rule out the acoustic-phonetic dimensions which are irrelevant to the perception of a nonnative contrast (in this case, English /r/-/l/ by native Japanese speakers), and in turn identify which dimensions are important for discriminating the contrast. 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 participated in a pretest phase, a training phase, and a posttest phase. Each phase involved a forced-choice identification task. The pretest and posttest were identical. The training phase contained 15 training sessions over 3 weeks. In each session, participants completed a forced-choice identification task with feedback regarding the correct answer after each trial. During each session, 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 that did not receive training did not show improvement. The amount of improvement varied depending on phonetic environment, with initial cluster and intervocalic environments improving more than initial singleton and final singleton environments, though final singleton position performance was near ceiling at pretest. Importantly, the authors concluded that this training procedure “was more robust than earlier training techniques” (1991, p. 874).

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 demonstrated that they were able to generalize their knowledge gained in the HVPT to the novel words produced by both talkers. In a second experiment, participants were trained using stimuli produced by only one talker. These participants showed improvement from pretest to posttest, but they were not able to generalize their learning to the stimuli produced by the new talker. The authors concluded that talker variability aids in robust category formation.

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 were retained long-term. They found that this new group improved from after training, and that participants performed significantly better on new words produced by a familiar talker than those from an unfamiliar talker. Results of a delayed posttest showed that 3 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 the posttest results and the delayed posttest results 6 months as suggesting that the high-variability training procedure can lead to long-term changes to in the representations of phonetic categories.

The high variability training protocol developed by Logan et al. (1991) came to be accepted by the field as one of the most effective laboratory speech perception training methods (Aliaga-García & Mora, 2009). High variability phonetic training, now commonly referred to as HVPT, established itself as a major field of inquiry pursued by researchers in the speech sciences. The following discussion highlights research and
findings on the topics of HVPT that are of particular relevance to language pedagogy, especially whether what is learned from HVPT can generalize to new listening contexts.

**HVPT and Second Language Acquisition**

In this section we discuss some methodological variations in HVPT perception and production research 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 to the application of HVPT to real-world L2 learning and teaching settings is the question of whether perceptual learning via HVPT can generalize to speech production, that is, more target-like pronunciation. Bradlow, Pisoni, Akahane-Yamada, and Tohkura (1997) studied whether HVPT that did not involve explicit pronunciation training could result in speech production gains. In their study, which consisted of 45 HVPT training sessions over 3 to 4 weeks, native Japanese speakers completed a pretest and posttest eliciting perception of the English /r/-/l/ contrast using a forced-choice identification task. Their production was also measured using 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 /r/-/l/ minimal pair identification task 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 3 months later. The authors interpret 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 holds promise for helping learners to improve their pronunciation of difficult L2 sounds.

Lambacher, Martens, Kakehi, Marasinghe, and Molholt (2005, experiment 2) investigated whether the production of five English vowels by native Japanese speakers learning English would improve following HVPT. The learners completed a pretest and posttest, during which they read aloud a list of twenty CVC minimal pairs. Between the tests, learners participated in six 20-minute HVPT sessions over 6 weeks. Twenty-six native English speaking judges listened to a subset of the productions and identified the vowel they heard in a five-alternative forced-choice task. The productions were also analyzed acoustically. Both the perceptual analysis by the native speakers and acoustic analyses indicated that learners’ pronunciations of English vowels improved following HVPT, whereas a control group of untrained participants demonstrated no improvement in either analysis. Summing up their findings, 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’ production.
The promising results obtained by Lambacher et al. (2005) were somewhat tempered by those obtained in Thomson and Derwing’s (2016) study. The researchers found that English learners from various L1 backgrounds who received 40 sessions of HVPT over a month showed differential improvement in production depending on both the type of stimuli used in the training and the task used to assess gains. Learners for whom the majority of the HVPT was made up of CV stimuli (which were often not real words) demonstrated small productive gains on English vowels in an elicited imitation task, while those whose HVPT contained mainly real-word stimuli did not improve. Neither group showed productive 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). More research is required to determine the extent to which HVPT can lead to pronunciation gains.

**Stimulus Variability**

As 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 represented in the training stimuli for optimal efficacy. As discussed above, 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); (three contexts rather than five). There was no effect of the number of phonetic environments on the size of perceptual gains in trained and untrained contexts 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 show evidence of generalization to new environments or talkers. The authors interpreted these data as indicating that single-talker training only results in “stimulus-specific learning rather than robust abstract category acquisition” (p. 13). In other words, while training that does not contain variability with respect to talkers can help listeners learn the particular words that they are trained on, it does not set learners up to develop L2 categories that are able to withstand the variability of the real world (i.e., the acoustic differences inherent to different voices and different contexts).

**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 of Mandarin, a tonal language. Specifically, the researchers investigated the extent to which the 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 multitalker training, while those with lower preexisting pitch perception ability benefitted more from singletalker training. For both groups of participants, there was better generalization to new talkers for those who had received multitalker training than with those who had only received single-talker training.

The authors followed up on these findings by looking more closely at the participants with lower preexisting pitch perception. They implemented three variations on the multitalker training procedures and found that when the multitalker training was
blocked by talker (i.e., listeners heard all the productions from talker A, then productions from talker B, then those from talker C, etc., rather than having the all three of the talkers’ productions mixed together), these learners were able to show greater gains. The authors interpreted these results as indicating that the high trial-by-trial variability (hearing a word produced by one talker and then immediately after hearing a new word produced by a new talker) was the source of the relative difficulty of HVPT for the low preexisting sensitivity group. The pedagogical takeaway from this research is twofold: (1) individual differences among learners may cause differences in the effectiveness of phonetic training, but (2) these differences can be mitigated if we understand the source of the difficulty. Blocking the HVPT by talker did not reduce the effectiveness of the multitalker training for those with high preexisting pitch perception, and it increased the effectiveness of the multitalker training for those with lower pitch perception. Thus, the authors suggest that this modification to the traditional HVPT method may result in the greatest overall gains for all learners.

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 the contributions of each to the efficacy of HVPT. 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 somewhat 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 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: (1) whether listeners are able to learn more than five L2 vowel categories at a time and (2) whether training on a few difficult vowel categories can result in improvement of other categories, as had been previously suggested with consonants. Specifically, they investigated whether training including a full L2 vowel set (in this case, American English monophthongs /ɪ, ɛ, æ, ɑ, ɔ, ʊ, u/) or only the most difficult subset of vowels for L1 Japanese speakers (/ɑ, ʌ, ʊ/) resulted in differential perception gains. Although participants trained on the subset of vowels showed greater gains on those three vowels than the participants trained on the full set of vowels, they did not demonstrate any gains on untrained vowels (i.e., 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 3 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 conclude 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 training emphasized phonetic information (that is, the majority of training stimuli were nonwords) or 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 a novel sentence containing each noun. The intelligibility of participants’ productions 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 primarily 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, they may make greater production gains.

**Synthetic enhancement of acoustic cues.** 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 (e.g., 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., F3 for /r/-/l/). While these studies have not revealed a benefit associated with modified stimuli relative to naturally-produced stimuli, it is worth noting that phonetically “enhanced” stimuli have been associated with increased phonetic category learning in other training paradigms (see, e.g., Escudero, Benders & Wanrooij, 2011).
HVPT Protocols

HVPT studies have also varied in terms of their overall duration and the type of training task used. Iverson and Evans (2009), 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 and discrimination) and found that, while the vowel perception improved after training using both tasks, participants trained using the identification task showed significantly greater improvement than those trained using the discrimination task. Hardison (2003) compared auditory–visual and auditory-only training, finding that audio-visual training led to significantly higher English /r/ and /l/ identification accuracy by native Japanese and native Korean 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 largely 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 agreement 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 phones. 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 2 decades and techniques that have actually been put into practice” (p. 540). To bridge this gap, their study involved some departures from earlier HVPT literature. First, the training took place over the course of 2 months, a longer period of time than most earlier studies. Second, participants had the flexibility to decide when, how often, and how many trainings they completed. These modifications were designed to simulate the option of extra practice outside of an ESL 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 2 months.

Despite these promising early results, very little other research has directly investigated the application of HVPT in nonlaboratory instructional 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 3 weeks in which they learned to associate vowels with images of nautical flags. Training stimuli were ten target vowels in two CV
contexts (/bV/ and /pV/) produced by 30 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, p. 752-753). Nineteen of the 22 participants demonstrated improvement in their pronunciation (measured using an elicited imitation task) following the training, even when imitating a new voice. Pronunciation improvement was shown in the trained CV contexts, as well as two of the four new contexts. Thomson concluded that “when designed using a principled, research-based approach, computer-mediated training…can improve speech intelligibility without explicit practice in production” (p. 758).

In a follow-up study (Thomson, 2012c), participants sampled from the same population participated in the same type of training and were tested to determine: (1) 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; (2) if perceptual improvements extended to new voices or new phonetic contexts; and (3) whether perceptual improvements were retained for a month after the training. Results did not demonstrate any differences between stimulus groups (enhanced, selected, or unselected vowels), but participants improve their vowel perception after training, even on productions by an untrained voice and on two of the four new contexts. Results from the delayed posttest showed that the improvement was maintained after a month but that there was no additional improvement, suggesting that the demonstrated improvement was indeed a result of the training.

Thomson 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” (2011, p. 760). To this end, Thomson created a freely available website called English Accent Coach that enables learners to choose which English sounds to work on (2012a). Training sessions using English Accent Coach are very similar the research studies described above. Tokens are produced by 30 different voices, target sounds are provided in a wide variety 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. Participants demonstrated productive gains following the computer-mediated training. Thomson (2012b) provides a useful list of ways to integrate English Accent Coach into the language classroom.

**Looking Forward**

The limited literature thus far provides evidence that the robust training effects of HVPT in laboratory studies can also occur in nonlaboratory settings and have pedagogical potential for both perception and production of L2 phones. In this section we discuss a number of lines of research that require attention in order for us 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 our understanding of 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 the efficacy of HVPT in the variety of language learning and teaching settings worldwide, more L1-L2 pairings need to be studied.

Generalizability of HVPT to New Phonetic Environments

Although effects of HVPT have been shown to generalize to new talkers (as reviewed above), it remains less clear how much 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 of the utility of HVPT for real-world language learners.

**Optimal Length of HVPT**

Given that HVPT studies have varied somewhat in their training stimuli and protocols, 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 where 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 applying 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 following 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 preexisting perceptual sensitivity of the listener to the trained contrast (e.g., Ingvalson, Barr, & Wong, 2013; Lee, Perrachione, Dees, & Wong, 2007; Perrachione et al., 2011). Lengeris and Hazan (2010) demonstrated that frequency discrimination performance on a
nonspeech continuum 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 continue 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 (see, e.g., Levis, 2016; Thomson, 2011), as it is low-cost and can be implemented via the internet (see, e.g., Thomson, 2012a). It is nonetheless necessary to empirically investigate these claims and the ultimate 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 relatively limited literature regarding the application of this training paradigm to real-world L2 learning and teaching settings suggest that HVPT is a tool with great potential to provide the input that is necessary for forming robust L2 phonetic categories with minimal extra effort on the part of the
teacher, and future research should focus on how to optimize the HVPT protocol for pronunciation learning.

References


Abstract

It has been well attested that infants and adults are able to take advantage of statistical distributional information to acquire phonemes (Hayes-Harb, 2007; Maye et al., 2002) and that infants can learn novel phonological alternations on the basis of phonological distributional information (White et al., 2008). Less is known about the way in which adult second language (L2) learners acquire allophonic relationships. The present study investigates the role of a phonological distributional mechanism in a controlled experimental context. We asked whether naïve subjects were able to utilize phonological distributional information to determine whether two phones belong to separate phonemes or a single phoneme. We exposed native English speakers to one of

---

2 Chapter 2 contains a complete manuscript, published as:
two artificial languages in which two acoustically similar sounds ([b] and [β]) occurred in either overlapping or complementary distribution. After the exposure phase, participants completed an ABX discrimination task. Unexpectedly, participants did not perform differentially on the task depending on their exposure type, failing to provide evidence for the use of a phonological distributional mechanism in adult L2 allographic acquisition.

**Introduction**

When acquiring the phonology of a language, learners must learn not only the contrastive sound categories, but also the positional variants (allophones) associated with those phonemes. Child language acquisition researchers have determined the timeline over which native language (L1) categories are acquired: vowels by 6 months (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 2006), most consonants by 8-10 months (Werker & Tees, 1984), and allophones by about 10-12 months of age (White, Peperkamp, Kirk, & Morgan, 2008). Several learning mechanisms have been identified to account for this rapid native phonological acquisition. Infant learners have been shown to be able to leverage a statistical distributional learning mechanism (Maye, Werker, & Gerken, 2002), as well as various lexical mechanisms, such as distinct visual referents (Yeung & Werker, 2009), and distinct word forms (Feldman, Myers, White, Griffiths, & Morgan, 2013) to support phonetic category learning. A phonological distributional mechanism has also been proposed for the learning of phonological status (Peperkamp, Le Calvez, Nadal, & Dupoux, 2006; Peperkamp, Pettinato, & Dupoux, 2003).

Research on adult L2 learners has also extensively investigated the acquisition of novel phonemic contrasts finding that numerous factors have an influence, including the
relationship between L1 and L2 phonetic and phonological systems, experience factors including age of acquisition and length of residence, and typological factors, among others (Best, 1995; Best & Tyler, 2007; Eckman, 1981; Flege, 1995). Evidence suggests that statistical distributional and lexical mechanisms may still be available for L2 phonological acquisition (Hayes-Harb, 2007). The present study investigates whether phonological distributional mechanisms, those thought to account for infant acquisition of allophone categories (Peperkamp et al., 2003; Peperkamp et al., 2006; White et al., 2008), may also be available for the acquisition of allophone categories by adult L2 learners.

**Background**

Traditionally phonological theory has made a distinction between two levels of representation—phonemes and allophones—which differ in their phonological distributions. Phonemes occur in overlapping distribution; they occur in the same phonological environments as other phonemes and are, therefore, not predictable by their context. Allophones occur in complementary distribution; different allophones of the same phoneme do not occur in the same phonological environments. Instead, the phonological context conditions which allophone occurs. Categories at each of these levels of representation differ cross-linguistically (e.g., [d] and [ð] map to a single phoneme category in Spanish but to two different categories in English), so both L1 and L2 learners must use information available in their input to learn them. The properties of the input required to do this are not yet determined, although various learning mechanisms have been proposed for native phonological acquisition.
Maye et al. (2002) provided evidence for a statistical distributional learning mechanism—the ability to track the frequency of particular phones in acoustic-phonetic space in order to develop phonemic categories. Lexical mechanisms—the ability to leverage distinct word forms, visual referents, and word-meaning pairs to bootstrap the formation of distinct phonetic categories—can also be used by infant language learners (Feldman et al., 2013; Yeung & Werker, 2009). Others have provided evidence that phonological distributional mechanisms—the ability to track phones with regard to their phonological environment—contribute to word segmentation, parsing, and allophone learning (Chambers, Onishi, & Fisher, 2003; Saffran, Aslin, & Newport, 1996).

Importantly, White et al. (2008) provided evidence that infants can use a phonological distributional learning mechanism to group allomorphic alternants into the same category.

Despite the differences that have been noted between L1 and L2 acquisition, evidence suggests that many of the learning mechanisms proposed for child language acquisition may also be available for and leveraged by adults. Hayes-Harb (2007) asked whether adult second language learners were able to use statistical and lexical information to learn a novel phonemic contrast. The results provided evidence that adult learners were able to use information about the acoustic-phonetic distribution of phones to learn novel contrasts. They were also able to use lexical information in the form of distinct visual referents to learn a novel contrast, suggesting that both statistical and lexical learning mechanisms continue to be available in second language phoneme category formation.

Peperkamp et al. (2003, Exp.2) investigated the role of phonological distributional
information (i.e., whether phones occurred in overlapping or complementary
distribution), in addition to statistical distributional information (i.e., whether phones
occurred in a monomodal or bimodal distribution along some phonetic dimension), on the
acquisition of phonemic and allophonic categories. Three groups of adult native French
speakers received exposure to an artificially manipulated continuum of the voiced and
voiceless uvular fricative [ʁ-χ], a French allophonic alternation, that varied with respect
to both the statistical distribution and the phonological distribution. Unexpectedly, the
interaction between exposure type and the difference in error rates on an AX
discrimination task at the pre- and posttest was not significant. However, only the group
exposed to the continuum in bimodal overlapping distribution showed a significant
difference between pre- and posttest. The authors interpreted these results as being in line
with their hypothesis that category formation is informed by both statistical information
and information about the contexts in which individual segments occur. However, the
lack of a significant interaction makes it difficult to draw a definitive conclusion. The
surprising results were attributed to a difference between the groups at the pretest, but
other factors may have also reduced the likelihood of observing an effect, such as the
variability in exposure tokens (i.e., 40 different words per target sound). Additionally, the
phones making up the target contrast were related as allophones of the same phoneme in
the participants’ native language, a factor known to reduce listener sensitivity to a
phonetic contrast (Boomershine, Hall, Hume, & Johnson, 2008). Thus, it may be that the
exposure phase in the experiment could not override the lifetime of exposure to the
alternation that the participants already had, or that any learning that could occur
happened during the pretest.
In another nonlexical perception study (i.e., one that does not require lexical access) investigating the acquisition of allophones by adult learners, Shea and Curtin (2010) asked whether learners are able to acquire knowledge of the relationship between L2 allophones and their phonological environments as they gain experience with the language. They compared four groups of listeners (monolingual English speakers, monolingual Spanish speakers, low-intermediate Spanish learners, high-intermediate Spanish learners) on a stress-detection task to determine whether more advanced learners performed more like native Spanish speakers with regard to expectations about the conditioning context of a Spanish allophonic alternant. In Spanish, voiced stops [b, d, g] alternate with approximants [β, ð, ɣ]; stops occur initially or after a nasal in the onset of a stressed syllable, whereas approximants occur intervocally and are posttonic. Like the native Spanish speakers, higher-proficiency Spanish learners showed a preference for associating the stressed syllable with initial stop consonants. Lower-proficiency learners and monolingual English speakers did not show this preference. The authors concluded that increasing experience with a second language results in increasing sensitivity to the factors that condition allophonic alternants and they attributed this learning to a phonological distributional learning mechanism. However, the learning of this association occurred prior to the experiment, failing to clarify which aspects of their input contributed to this acquisition.

Taken together, although it is compelling to claim that a phonological distributional mechanism remains available for L2 phonological acquisition, no study to date has provided clear evidence that this mechanism is still available for use in adult allophonic acquisition. We attempt to build on this early literature by using participants
with relatively little experience, controlling their input, and isolating phonological
distributional information as an independent variable.

The Present Study

We investigate whether adult second language learners can use phonological
distributional information (in the absence of lexical-semantic information or explicit
instruction) to determine whether two phones belong to the same or different underlying
categories. To this end, we expose relatively inexperienced listeners to a novel contrast,
such that the two target phones occur in either overlapping or complementary
distribution. We predict that learners trained on the contrast in overlapping distribution
will infer two categories (a phonemic distinction), whereas learners trained on the
contrast in complementary distribution will infer that the two phones are related as
allophones of a single underlying phoneme category. Perceptually, adult native speakers
are less sensitive to the phonetic distinction between allophones of the same phoneme
(Boomershine et al., 2008; Whalen, Best, & Irwin, 1997) than to allophones of separate
phoneme categories, so we expect participants in the complementary exposure group
category to perform less accurately on an ABX task than those in the overlapping
exposure group. The ABX task was selected because it can be completed by participants
who are naïve with respect to the tested language, yet it is typically assumed to tap into a
more phonological level of representation than does the AX task used by Peperkamp et
al. (2007).
Participants

Participants were 40 native English speakers (M = 12 F = 27 Other gender = 1) between the age of 17-52 years old (mean = 24.15) recruited at the University of Utah from the Salt Lake City area. Participants had studied Spanish for no more than 1 year, and had no history of hearing, developmental, or neurological disorders. They were randomly assigned to one of two exposure groups—Complementary group (n = 20, M = 8 F = 11 Other gender = 1, mean age = 23.5) or Overlapping group (n = 20, M = 4 F = 16, mean age = 24.6). Participants were either compensated $10 for their time or received course credit for their participation.

Stimuli

The stimuli were natural recordings produced by a Spanish-English bilingual with phonetic training. Each nonword stimulus was produced in the carrier phrase Diga la/una __________, por favor (Say the/a __________, please). An acoustic analysis of the intensity ratios, a measure of degree of consonant constriction, was made for each production of the target contrast [b-β] following Carrasco, Hualde, and Simonet (2012) to ensure there was indeed an acoustic distinction to learn (mean [β] intensity ratio = 82.26, mean [b] intensity ratio = 60.08, consistent with greater constriction for [b] than [β]). Four tokens of each stimulus type were selected for the task. The target sounds occurred in a syllable with [a] and the target syllable was either preceded or followed by [ti] and [ku]. While this is much less variability than typical of the real world, this was a conscious choice to reduce variability in other syllable types, given that Peperkamp et al. (2003) suggested the variability of their exposure tokens may have introduced unwanted
noise in their study. Stress was controlled such that all stimuli had penultimate stress. The overlapping exposure condition contained [b] and [β] in both initial and medial position, whereas in the complementary condition [b] was restricted to initial position and [β] to medial position. Because the complementary exposure tokens were a subset of those used for overlapping exposure, each of the target nonwords in the complementary exposure condition was played twice as frequently to ensure both groups receive exposure phases of equal length. Control stimuli were identical to the target nonwords except the target contrast was replaced with a native contrast, [m-l]. The distribution of [m] and [l] was not constrained for either exposure group; both groups were exposed to tokens containing [m] and [l] in both initial and medial position. A summary of the stimuli is provided in Table 2.1.

**Procedure**

The experiment consisted of two parts: an exposure phase and a test phase.

**Exposure Phase**

During the exposure phase participants heard 768 auditory tokens (12 randomized blocks x 4 tokens x 16 words) separated by an ISI of 500 ms. Participants were instructed to listen and check a box on the sheet provided each time they heard a word. The exposure lasted approximately 16 minutes.
Table 2.1 Stimuli heard in each exposure group.

<table>
<thead>
<tr>
<th>Complementary</th>
<th>Overlapping</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>initial</td>
</tr>
<tr>
<td>Experimental contrast</td>
<td>[b]</td>
</tr>
<tr>
<td></td>
<td>[β]</td>
</tr>
<tr>
<td>Control contrast</td>
<td>[m]</td>
</tr>
<tr>
<td></td>
<td>[l]</td>
</tr>
</tbody>
</table>

Test Phase

The test phase consisted of an ABX discrimination task in which the participant indicated whether the third sound was the same as the first or the second. Each token (4 tokens x 16 words = 64 trials) appeared as X; in each trial, the three words were all physically different tokens. Trials were counterbalanced for A and B matches and were presented in a random order for each participant. The test was divided into two blocks of 32 trials. Test items were separated by an ISI of 500 ms and trials timed out at 2500 ms.³

³ Although the complementary distribution group were exposed to only [b] initially and [β] medially during the exposure phase, due to the nature of an ABX task, during the test phase they received input to each target token in the “wrong” position. For example, in trial [bati] – [βati] – [βati], participants were exposed to two instances of initial [β]. We ensured that the ratio of “wrong” test items to “right” training items remained low (9%). This concern informed the duration of the exposure phase as well as the number of trials in the test phase.
Results

Two analyses were conducted. For the first, the dependent measure was mean proportion correct, and for the second, the dependent measure was d-prime score. As seen in Figure 2.1, the mean proportion correct for the complementary exposure group was 0.85 for the target contrast and 0.88 for the control contrast. For the overlapping exposure group, mean proportion correct for the target contrast was 0.84 and for the control contrast was 0.85. A two-factor mixed-design ANOVA was conducted with exposure group as the between-subjects factor (two levels: complementary and overlapping) and contrast as the within-subjects factor (two levels: [b- β] and [m-l]). There was no significant main effect of exposure group $F(1, 38) = 0.498, p = 0.484$ or of contrast $F(1, 38) = 2.791, p = 0.103$. Crucially, there was no significant interaction between group and contrast $F(1, 38) = 0.112, p = 0.740$, indicating that the two groups did not differ in their mean proportion correct across the contrasts.
differ from one another in their performance on the [b-β] contrast.

In addition to considering the mean proportion correct, d’ scores were also computed and submitted to statistical analysis. D-prime is a measure of sensitivity that factors out individual response bias. As seen in Figure 2.2, the mean d’ score for the complementary group was 2.22 for the target contrast and 2.39 for the control contrast. For the overlapping group, mean d’ score for the target contrast was 2.07 and for the control contrast was 2.18. A second two-factor mixed-design ANOVA was performed. Again, there was no significant main effect of exposure group $F(1, 38) = 0.698, p = 0.409$ or of contrast $F(1, 38) = 2.754, p = 0.105$, and there was no interaction of group and contrast $F(1, 38) = 0.140, p = 0.710$.

*Figure 2.2 Boxplot of d-prime scores by exposure group and contrast. Superimposed text represent group means.*
Discussion

This study investigated the role of phonological distributional information in adult L2 phonological acquisition. We asked whether participants exposed to a novel phonetic contrast [b-β] in either complementary or overlapping phonological distribution performed differentially on an ABX task following the exposure. Unexpectedly, we observed no difference in the two groups’ performance on the [b-β] contrast, suggesting that participants were not able to use phonological distributional information alone to determine whether the two phones were allophones of the same or different phonemes. These results are consistent with those of Peperkamp et al. (2003), suggesting that perhaps, at least under these constrained laboratory conditions, a phonological distributional mechanism alone may be insufficient for adult learners to infer the phonological status of a novel pair of phones.

We attempted to build on earlier work by controlling for previous input regarding the novel phones, using a measure of perception that relies on higher levels of phonological representation, and isolating the variable of phonological distributional information as a potential contributor to the acquisition of a novel allophonic alternation. Previous research (i.e., Peperkamp et al., 2003 and Shea & Curtin, 2010) used stimuli created by splicing allophones produced in their conditioned contexts into phonotactically illegal contexts, while the stimuli in this study were naturally produced in illegal contexts. This solution may have eliminated extraneous cues introduced by cross-splicing. However, it also had limitations, such as the difficulty of producing phones in unnatural positions and the necessity of ensuring that productions in illegal contexts were produced with similar intensity ratios as those in legal contexts. The stimuli were chosen
based on a phonetic analysis of the intensity ratio between the target consonant and the following vowel [a] as well as the presence/absence of a burst. Carrasco et al. (2012) showed that there is a bimodal distribution of the intensity ratios of [b] and [β] in both Madrid and Costa Rican Spanish, so we sought to verify that a similar distribution was present in our stimuli. However, some research suggests that allophones are not only perceived as more similar than contrastive phones, but that they are also produced more similarly in languages for which they are allophonic than they are in languages for which they are phonemic (see e.g., Seidl & Cristia, 2012). It is possible that we artificially enhanced the statistical distribution of these phones due to this analysis.

The task used in this experiment was carefully considered, as the foundational literature in L2 allophone acquisition is relatively sparse. Shea and Curtin (2010) successfully showed gains in allophonic knowledge of learners using a metalinguistic task that demonstrated understanding of allophones with respect to their conditioning contexts. Because we asked a different question—whether learners could make use of phonological distributional information to group phones as allophones of a single phoneme category—we chose a different task. Using a task that relied on the perceptual correlates of native phonemes and allophones as a measure was a less metalinguistic way of detecting early knowledge about phonological status. This task also had limitations, such as the necessity of including tokens that one of the groups did not hear during the exposure phase in the test and the possibility that phonological status could be “unlearned” through this conflicting evidence. It is also possible that participants used an acoustic strategy, such as treating B and X as an AX task and ignoring A altogether, a possibility that could be effectively eliminated by using more than one talker to produce
stimuli in future research.

At the same time, the null effects may have been caused by an unexpected pattern of perceptual assimilation. The alternation was chosen to be maximally similar to Shea and Curtin (2010), who looked at the entire natural class that undergoes this alternation in Spanish, /b,d,g/. We considered the alternation [g]-[ɣ] because it would likely result in a same-category assimilation in English (Best, 1995), but we were concerned about the perceptual salience of this distinction. Additionally, an acoustic analysis of intensity ratios found that these phones are monomodally, as opposed to bimodally distributed in Spanish (Carrasco et al., 2012). We avoided the third pair [d]-[ð] because they are contrastive in English. English does not have a category /β/, so the study was designed as if participants would hear one familiar sound and another that was also assimilated to /b/ but may have been a poorer exemplar of the category. However, pilot studies indicated that [β] was frequently assimilated to the category /v/ by English speakers. It could be that this two-category assimilation (Best, 1995) blocked distributional learning during the short exposure period. While we strategically chose a pair that contained one familiar and one unfamiliar member for English speakers, it is worth noting that the L1 English L2 learner of Spanish would be required to master all of these assimilation patterns.

Moreover, although earlier studies have shown that infants can learn from distributional information with as few as 2 minutes of exposure (Saffran et al., 1996) and adults with as few as 9 (Hayes-Harb, 2007; Peperkamp et al., 2003), it is also possible that a longer exposure period may be required to learn this particular alternation or allophonic alternations in general.

Finally, we would have ideally used participants who were completely naïve to
Spanish, but it proved challenging to find participants without Spanish experience in the Salt Lake City area. It is possible that informal and/or limited formal exposure to Spanish (which was not controlled for in this study) provided prior input from which participants could have learned about the alternation.

This study attempted to tightly control many variables that researchers noted may have contributed statistical noise in their studies and masked small but reliable effects from being detected. Despite these modifications, we failed to find support for the hypothesis that adult second language learners are able to use phonological distributional information alone to learn that two phones are related as allophones of the same phoneme. Nevertheless, there is much to build upon. For future research, other ways of selecting stimulus tokens should be considered and other contrasts should be investigated. Alternations from other languages may be better for recruitment purposes. Despite the null results, the role of phonological distributional information is an important area of research in which study should continue in order to deepen our understanding of adult second language phonological acquisition.

References


Shea, C. E., & Curtin, S. (2010). Discovering the relationship between context and


CHAPTER 3

HIGH VARIABILITY PHONETIC TRAINING AND WORD LEARNING:
AN INVESTIGATION OF THE GENERALIZABILITY OF
PERCEPTION GAINS TO HIGHER-LEVEL TASKS
FOR SECOND LANGUAGE LEARNERS

Introduction

Since Logan, Lively, and Pisoni’s (1991) seminal study exploring the role of phonetic variability in the acquisition of second language (L2) phonology, speech research has repeatedly affirmed the effectiveness of a training model called high variability phonetic training (HVPT) for improving perception and production of nonnative contrasts in a laboratory setting (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Carlet, 2017; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Lively, Logan, & Pisoni, 1993; Wang & Munro, 2004). Indeed, it is considered to be one of the most effective laboratory speech training paradigms (Aliaga-Garcia & Mora, 2009). More recently, HVPT has also been shown to lead to perception and production gains on discrimination and identification tasks when implemented outside of a laboratory and by real L2 learners (Iverson & Evans, 2009; Nishi & Kewley-Port, 2007; Thomson, 2011; Thomson, 2012). Increasingly, L2 pronunciation researchers have called for more research to bridge the gap from theoretical findings to instructional practice,
including determining whether the perception gains from HVPT are robust to less controlled situations, such as part of a class or as extra at-home practice.

Some evidence suggests that laboratory phonetic training can also improve word learning performance (Ingvalson, Barr, & Wong, 2013). This is notable because, as Curtin, Goad, and Pater (1998) point out, learners may not be able to make use of features necessary for L2 lexical representations if they are not “present lexically in the L1, even though they may be able to discriminate [the relevant contrasts] on the basis of surface features” (p. 10). They affirm the necessity of treating nonlexical and lexical tasks as distinct in developmental speech perception research. Ingvalson et al. (2013) found, however, that gains from perceptual training may generalize to lexical tasks. To the extent that such generalization is possible, the question of the time course of the development of perception and the ability to generalize that knowledge to lexical tasks remains. These abilities may develop in parallel, such that as perceptual knowledge is updated, it can immediately be implemented in lexical tasks; alternatively, they may develop independently, such that successful implementation in lexical tasks lags behind perceptual task improvement (but see Darcy et al., 2012).

**Background**

Speech research indicates that humans are born with the capacity to perceive all language sounds they have been tested on. With exposure to the language that will become their native language over the first year, infants’ perceptual systems hone in on the contrasts that are meaningful in that language and effectively lose sensitivity to irrelevant contrasts (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker &
Tees, 1984;). As a result, adults are sensitive to the language-specific phonemic contrasts of their L1 but are less sensitive to noncontrastive phones (Boomershine, Hall, Hume, & Johnson, 2008). Models of cross-language and L2 speech perception (e.g., the Native Language Magnet Model, Kuhl, 1993; the Perceptual Assimilation Model, Best, 1994; the Speech Learning Model, Flege, 1995) attempt to explain why an adult’s established perceptual system may or may not be sensitive to particular nonnative sounds. These models largely focus on the relationship between the L1 and L2 phonological inventories for the relative difficulty of learning an L2 category. In many cases, target-like perception of a novel L2 phoneme can be very difficult. Despite these challenges, people are able to learn new L2 contrasts, and phonetic training is one technique that can help us understand how this happens.

**Phonetic Training**

Phonetic training is “an effort to change patterns of speech perception in learners through relatively short-term laboratory training methods” (Burnham, 2013, p. 89). There are different models of phonetic training (e.g., lexical as in Hayes-Harb, 2007 and Strange & Dittman, 1984; distributional as in Hayes-Harb, 2007 and Maye, Werker, & Gerken, 2002), that each focus on the contribution of a particular feature of the input to phonological acquisition. This paper will focus on the model that has received the most attention in the field over the last few decades: high variability phonetic training (HVPT).

HVPT was originally designed to investigate the role of variability in phonological acquisition. Logan, Lively, and Pisoni (1991) hypothesized that the types of variability that naturally occur in the production of sounds by native speakers, such as
different voices and different phonetic environments, may help learners narrow their focus to the phonetic features that crucially differentiate a nonnative contrast. To test this hypothesis, Japanese speakers completed a forced-choice identification training in which they heard a word and had to select which member of the minimal pair they heard. If the wrong member of the pair was chosen, the auditory form was repeated and the correct answer was highlighted. The training stimuli were 68 minimal pairs contrasting the difficult English phones /l/ and /ɹ/ produced by six different talkers. For the pretest and posttest, participants heard 16 of the trained words and selected which member of the pair they heard. After training, participants also completed two tests of generalization containing words that were not part of the training as produced by either a talker from the training or a new talker. Participants demonstrated higher accuracy on the final identification task showed greater perceptual improvement and were able to generalize this knowledge to a new talker. Lively, Logan, and Pisoni (1993) followed up on this finding by replicating the results and adding a control group trained only on stimuli produced by a single talker. Results showed that participants who were trained using HVPT outperformed those who were trained on stimuli produced by only a single talker or in a single context. Participants who had received HVPT were also more accurate on trials with new talkers.

Based on these findings and the many replications that followed, the field of L2 speech research has concluded that “input that fully represents the range of variation in normal speech leads to more robust acquisition of an L2 sound system” (Levis, 2016). As such, research employing HVPT over the last decade has focused not on whether variability contributes to L2 phonological acquisition, but the extent to which this
training protocol can contribute to language acquisition in both laboratory and applied settings.

**Lexical Learning**

The majority of HVPT research continues to focus on what might be called “low-level” perception tasks (i.e., identification and discrimination) as the measure of linguistic development, but some researchers argue that “knowledge of a phonetic distinction is useful linguistically only if it is exploited to distinguish words in the language” (Hayes-Harb, 2007).

To understand the utility of HVPT in language acquisition, it is thus reasonable to ask whether HVPT can support subsequent word learning ability. Ingvalson et al. (2013) provided some preliminary evidence that HVPT can be helpful for word learning in a laboratory setting. Native English speakers learned minimal triplets of pseudo-words distinguished only by lexical pitch in one of two conditions. In the first condition, participants received 3 days of HVPT on the three pitch patterns and then 5 days of lexical training. In the second, participants only received 8 days of lexical training. During the HVPT, participants heard a syllable with a superimposed pitch, selected which of three arrows depicted the direction of the pitch contour, and received corrective feedback. During the lexical training, participants heard the same stimuli, chose one of three pictures that depicted the meaning of the pseudowords, and received corrective feedback. At the end of each day of lexical training, listeners completed a final identification task in which they heard a pseudoword, chose from the full set of 18 possible meanings, and did not receive feedback. Participants in the HVPT condition
performed more accurately on the final vocabulary test than those in the lexical-only condition. Replication of these findings using another kind of artificial lexicon task (see e.g., Escudero, Hayes-Harb, & Mitterer, 2008; Hayes-Harb, 2007) with less lexical training, together with Ingvalson et al. (2013), would provide more robust evidence for the relationship between HVPT and subsequent word learning ability; as noted by Sakai and Moorman (2018) “researchers should never rely on a single study to answer large and theoretically important questions” (p. 191). It is also important to seek evidence that HVPT may generalize to word learning for real-world L2 learners, as this initial study involved naïve participants.

**HVPT as a Pedagogical Tool**

Traditional phonetic training studies are conducted with participants in tightly controlled phonetics laboratories, often focusing on target contrasts from languages that the participants are completely naïve to. In the last ten years, HVPT research has moved toward being more learner-centered (see e.g., Iverson & Evans, 2009; Thomson, 2012) by conducting studies in less controlled environments and using real L2 learners as participants. However, current research on the benefits of HVPT remain underrepresented in teaching materials (Levis, 2016). This disconnect is especially apparent when one takes into consideration the rapidly expanding fields of Computer Assisted Language Learning and Computer Assisted Pronunciation Training (CAPT; see, e.g., Levis, 2007 for an overview). Many laboratory phonetic training studies have used a computer-mediated approach to focus attention on the input and maximize the effect of exposure (Lively et al., 1993; Logan et al., 1991; Pisoni & Lively, 1995; Wang & Munro, 2004).
HVPT via computer-mediation has also been used in real learner programs and has demonstrated results in improved perception (Iverson & Evans, 2009; Nishi & Kewley-Port, 2008) as well as production (Thomson & Derwing, 2016). Thomson (2011) claimed that computer-mediated HVPT can provide practice that is not always practical during regular classroom instruction and may be an effective supplement to other instructional practices for pronunciation improvement.

To this end, English Accent Coach (EAC; Thomson, 2012b) was created to be a low-resource program that can easily be integrated into current English as a second language (ESL) programs, even serving as homework. This program is a freely available online HVPT platform with the capability to train all English consonants and 10 English vowels. The auditory stimuli are produced by 30 Canadian English speakers in a variety of phonetic environments. Learners can select which consonants or vowels to be included in the training. Like the laboratory studies, the training is a forced-choice identification task, in which participants hear a stimulus and select the International Phonetic Alphabet (IPA) symbol that represents the sound they heard. Each trial contains corrective feedback in the form of a sound (ding for correct and buzz for incorrect) and color (green for correct and red for incorrect), and the correct answer must be selected before the program moves to the next trial. This program was designed as “an extension to…research-based pedagogy” for the purpose of “furthering large-scale research into the efficacy of HVPT” (Thomson, 2012, p. 1254). As such, it is an excellent candidate for bridging the gap between the traditional laboratory studies and more learner-centered applications.

If HVPT results in greater ability to learn new words for real language learners,
this would suggest that learning from this type of training may be able generalize to other kinds of linguistic tasks. In this case, HVPT would have wider appeal for real language programs even outside the area of pronunciation (the area in which it has been most widely studied). If, on the other hand, HVPT is not able to generalize to learning new words, this may suggest that other types of L2 exposure may be needed for learners to make lexical use of their knowledge of L2 contrasts.

The Present Studies

In two experiments, we investigate whether real ESL learners can exploit knowledge gained from an online implementation of HVPT to establish and use lexical representations that encode difficult L2 phonological contrasts. Mandarin speakers from two different populations received HVPT of English vowels using English Accent Coach either as part of an ESL class or on their own time using computers available to them. Preceding and following training, they completed a discrimination task, which was expected to demonstrate improvement in their perceptual abilities with respect to three difficult English vowel contrasts after the training. Preceding and following training, they also participated in a word learning task which examined their ability generalize any gains in perception of the three contrasts to lexical encoding and retrieval.

Experiment 1

This first experiment (E1) aims to contribute to the HVPT literature by providing an initial attempt to apply this training protocol in an ecologically valid setting—that of an already-existing ESL instructional situation. As such, it differs from earlier studies in a
number of ways, described here.

First, the study was conducted using existing groups of students rather than randomly assigning participants to conditions and was thus a pseudo-experiment. This means that that there may be important differences between groups that are not of immediate relevance to the research questions, including general class affect, the time and day of the class, and the relationship between the instructor and the class. These are discussed in more detail below.

Second, the training was done within the constraints of class meetings rather than in a laboratory or on their own time. This means that participants were restricted to the time and location in which the class was offered; thus enforcing completion of a specified number of training sessions was not practical. Rather, students completed the amount of training they could in the time available, and this varied by student.

Third, participants completed the testing during class time as well, and in a classroom setting, whereas in previous studies participants were tested in laboratory settings (e.g., Lively et al., 1993; Logan et al., 1991). This means participants were in a room with many other participants at a time, both making it more difficult to ensure that each participant understood the task and providing more opportunity for distraction during the testing session.

Finally, learning was operationalized somewhat differently than in earlier work on HVPT. First, perception gains were assessed via a discrimination task rather than an identification task (as is more typical in earlier work, see e.g., Logan et al., 1991; Nishi & Kewley-Port, 2007; Thomson, 2012), although the HVPT program utilizes an identification task for training. There are two reasons for this, both of which have to do
with our inclusion of an untrained control group (which most previous studies do not include). First, the training itself involved an identification task, which means that over the course of the experiment the trained group accumulated experience with an identification task that the control group did not. In addition, the training involved IPA symbols and thus served to familiarize the trained participants with the use of these symbols. Because an identification task typically requires labels for the sounds, and because English spelling does not provide for a one-to-one mapping between vowel sounds and letters, we determined that the optimal option for collecting identification data is to use IPA symbols. Given that trained participants were exposed to the IPA over the course of the study but control participants were not, the use of IPA symbols in an identification task at test would have resulted in a confound in the study design.

Another way in which the operationalization of perceptual learning in the present study differs from most previous work in this area is that we also used a word learning task at test. The purpose of the word learning task was to examine participants’ ability to generalize perceptual (nonlexical) learning to a higher level (lexical) task. Wong and Perrachione (2007, p. 566) suggest that there is “a phonetic-phonological-lexical continuity for speech learning, such that more basic auditory abilities (phoneme discrimination) mediate performance on higher level auditory tasks (word learning)” and that a full understanding of language processing and instruction requires understanding the relationship between the acquisition of sound categories and words.
Participants

Participants were recruited from a group of sports coaches at Chinese universities who were at the time participating in a 3-month coaching education program at the University of Utah. Part of the coaching education program included 5 hours of English instruction per week, and it was during these class meetings that they participated in the present study. For all of them, it was their first time in an English-speaking country, and they had novice-low English proficiency according to the ACTFL proficiency guidelines. All participants planned to return to China at the end of the program to continue their coaching careers at Chinese universities.

The coaches were assigned to classes by sport, and we had no a priori reason to believe that coaches of different sports would have systematically different English language proficiencies or motivation to learn English. These classes were randomly assigned to two conditions: HVPT via English Accent Coach (experimental) or No Training (control). While all 75 coaches were invited and encouraged to take part in the experiment, many opted out of participation, failed to complete one or both parts of the pretest and/or the posttest, or timed out of over 50% of one or both parts of either the pretest or the posttest. In the end, only 15 coaches completed all four pre- and posttest tasks, 12 of whom had completed some amount of training and three of whom had not. Self-reported demographic data for the 28 coaches who elected to participate are provided in Table 3.1.
Table 3.1 Self-reported demographic data from the 28 coaches who elected to participate.

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Age</th>
<th>Age at first classroom experience with English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men’s basketball</td>
<td>9</td>
<td>1</td>
<td>36-45</td>
<td>9-13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>m = 38.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD = 2.6</td>
<td></td>
</tr>
<tr>
<td>Women’s basketball</td>
<td>8</td>
<td>0</td>
<td>31-44</td>
<td>12-23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>m = 37.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD = 4.6</td>
<td></td>
</tr>
<tr>
<td>Track and Field</td>
<td>9</td>
<td>1</td>
<td>36-44</td>
<td>13-24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>m = 38.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD = 3.6</td>
<td></td>
</tr>
</tbody>
</table>

Training Stimuli and Methods

English Accent Coach (EAC; Thomson, 2012b) is a freely available web-based perceptual training program with the capability to train all English consonants and 10 English vowels. A researcher mode of EAC is available such that training sessions can be precisely controlled. In order to determine which three vowel contrasts to train the Mandarin speakers on, a pilot AXB study was conducted. Participants were seven native speakers of Mandarin recruited from the University of Utah community. One participant’s data were not included because he failed to respond in 40% of the trials. Stimuli were produced by three talkers whose voices are included in the training voices on English Accent Coach. Consonants were /b, p, d, t, g, z, s, v, f, j/. Six vowel pairs were identified from the literature (Jia, Strange, Wu, Collado, & Guan, 2005; Lai, 2010; Thomson, 2007, 2012; Wang, 1997) as being difficult for native Mandarin speakers: /i-i/, /i-e/, /e- æ/, /æ- a/, /a-a/, and /ʌ-ʊ/. Target items were productions of CV syllables containing every combination of consonant and vowel (77 tokens produced by three speakers), totaling 231 tokens. In the AXB task, A, X, and B were each produced by a different talker and there was one trial per CV syllable per contrast (e.g., there was one
trial bi-bi-bi and one trial bi-bi-be). The match (AX or XB) was counterbalanced for each contrast. At least four of the six Mandarin speakers had lower than 70% accuracy on /ɪ-ɛ/, /ɛ-æ/, and /α-ʌ/. However, the lowest overall accuracy was on /ɛ-æ/ (64%), /ʌ-ʊ/ (64%) and /α-ʌ/ (58%). Based on these two measures, /ɛ-æ/ and /α-ʌ/ were identified as two difficult contrasts. To determine whether to use /ɪ-ɛ/ or /ʌ-ʊ/ for the third contrast we considered the following: (1) The standard deviation (SD) for /ʌ-ʊ/ was 0.19, and the SD for /ɪ-ɛ/ was 0.10; (2) In the literature, only one study was identified that looked at the /ʌ-ʊ/ contrast, whereas two different confusion matrices identified /ɪ-ɛ/ as a troublesome contrast for Mandarin speakers. The only confusion matrix that included both of these contrasts indicated that /ɪ-ɛ/ were more often confused than /ʌ-ʊ/. (3) /ʊ/ is a relatively low-frequency sound (Brown, 1988), so it is unlikely to inhibit communication. For these reasons, we selected /ɪ-ɛ/, /ɛ-æ/, and /α-ʌ/ as the three contrasts to focus on.

In this study, the HVPT group received perceptual training for five English vowels /ɪ, ɛ, æ, ʌ, α/, as these make up the three pairs of vowels that were most perceptually difficult for the native Mandarin speakers in the pilot task. The vowels were trained in the context of CV syllables (all phonotactically legal English consonants were used) produced by 30 speakers of Canadian English. During each training session, participants responded to 200 randomly presented auditory stimuli by selecting the phonetic symbol of the vowel in the word they heard from a visual vowel chart, as shown below in Figure 3.1. Following each selection, the program provided both auditory (a “ding” for correct and a “buzz” for incorrect) and visual (the selected symbol was highlighted green for correct and red for incorrect) feedback. All five vowels were trained in each session.
Participants were provided with opportunities to participate in the training in a computer lab during their class time for 3 weeks. Instructors were told that the goal was for each participant to complete 18 training sessions. A number of unforeseen events over the course of the 3 weeks, including difficulties logging on to the computers and a tragedy on campus that resulted in university-wide class cancellation, resulted in much lower amounts of training than expected. The 17 experimental participants whose data were analyzed completed between five and 13 sessions over the course of 3 weeks. The seven control participants did not receive training.
Testing Stimuli and Methods

The pre- and posttests were conducted in a computer lab during class time. The experimenter provided instructions for the whole class at once. Although the instructions were identical for both groups, only the HVPT group had access to an interpreter to help us answer their questions during the pretest. At the pretest, participants provided informed consent and filled out a background questionnaire containing questions about their language and cultural background as well as their experience with English. Participants were assured that their participation was voluntary. The pre- and posttest tasks were identical in all ways except for stimulus items. Each test contained two parts: a word learning task and a discrimination task.

The word learning task. The word learning (WL) task included both auditory and visual stimuli. The auditory stimuli, in Table 3.2, consisted of CVC English-like nonwords containing the five vowels that the HVPT group was trained on. Initial consonants were /b, d, t, g, k, h, f, s, z, j/ and final consonants were /b, p, d, g, k, f, v, s, z, j/. Stimuli were produced by three native speakers of North American English whose voices were not included in the training stimuli. The native English speakers read the stimuli in sentences that contained rhyming words to help them with pronunciation (e.g., fish rhymes with kish). The sentences were produced by each talker six times and three were selected as stimuli for the task. Within a single word learning task (either pretest or posttest), minimal pairs were not taught.

4 Many thanks to all the volunteers who helped proctor the experiment sessions: Lauren Brocious, Samantha Dean, Rachel Hayes-Harb, Rachel Haynes, Mary Lamb, Lisa Leatherwood, Joselyn Rodriguez, Eric Schoening, Catherine Showalter, David Thomas, Heather Thomas, Sue Veach, and Julia Vonessen. Thanks also to Jack for translating.
Table 3.2 Word learning stimulus items

<table>
<thead>
<tr>
<th>Vowel Pairing</th>
<th>Pretest Stimulus Items</th>
<th>Posttest Stimulus Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ɪ-ɛ/</td>
<td>kɪʃ</td>
<td>kɛʃ</td>
</tr>
<tr>
<td></td>
<td>hɛs</td>
<td>hɛs</td>
</tr>
<tr>
<td>/ɛ- æ/</td>
<td>fɛk</td>
<td>fɛk</td>
</tr>
<tr>
<td></td>
<td>fɛv</td>
<td>fɛv</td>
</tr>
<tr>
<td>/ɑ-ʌ/</td>
<td>sɑɡ</td>
<td>sɑɡ</td>
</tr>
<tr>
<td></td>
<td>dɑs</td>
<td>dɑs</td>
</tr>
</tbody>
</table>

Visual stimuli were line-drawings of nonobjects, such that each of the 24 nonwords was assigned a “meaning” represented by a nonobject picture to ensure that the participants did not already have a word in their lexicon for the provided pictures.

The word learning task was made up of three phases: word-learning, criterion test, and final test (see Figure 3.2). During the word learning phase, participants heard a nonword and saw its associated nonobject picture. Participants were told to learn the words and their meanings, but no response was required. Each word was presented two times per block in four blocks (eight times total, produced by each talker two or three times), and each block was presented in a different random order for each participant.

During the criterion test, participants saw a nonobject and heard a nonword. Each image was presented four times, twice in a matched condition and twice in a mismatched condition. Mismatched auditory forms were one of the other nonwords they heard during the word learning phase containing a vowel from a pair that is different from that of the target form (e.g., if the matched auditory form was [kɪʃ] the mismatched form might be
Figure 3.2 The three phases of the word learning task. Eyes represent what the participant saw, ears represent what the participant heard, and the hand indicates the required response. Phase 1 required no response.

Participants responded to each visual-auditory pair by pressing YES (matched) or NO (mismatched). Participants repeated the word-learning criterion test as many times as necessary to reach 90% accuracy on the criterion test. The final test was the same as the criterion test except that the auditory form in the mismatched condition formed a minimal pair with the target form (e.g., if the matched auditory form was [kɪʃ] the mismatched form was [kɛʃ]).

The discrimination task. The second part of both the pretest and posttest was an AXB discrimination task (AXB), in which participants heard were asked to indicate whether the second nonword in a triad was the same as the first or the third. The auditory stimuli, provided in Table 3.3, consisted of both CV syllables (that were included in the training) and CVC English-like nonwords (not included in the training, but included in the word learning task) containing the five vowels that the HVPT group was trained on. They were produced by the same three native speakers of North American English who
produced the stimuli for the word learning task. Talkers read the stimuli in sentences that contained rhyming words (e.g., Fish rhymes with kish) or words that started with the CV syllable (e.g., Zombie starts with zah) to help them with pronunciation. The sentences were produced by each talker six times and three were selected as stimuli for the task.

For each trial, A, X, and B were each produced by a different talker. Each contrast was tested 12 times per block with four blocks such that each talker produced each of the four words four times as X, and talker order, as well as match (AX or XB), was counterbalanced within the blocks.

### Analysis and Results

The discrimination task and the word learning task were considered separately due to the small number of participants who completed both tasks. Data were analyzed from participants who: a) completed both the pretest and posttest of the same task, and
either b) timed out for no more than five trials in a row for the AXB task or c) timed out for no more than two trials in a row for the WL task. Table 3.4 indicates the number of people in each class who completed each task. Twenty-four participants completed both the discrimination pre- and posttests (HVPT group \( n = 17 \), No Training group \( n = 7 \)) and 15 participants completed both the word learning pre- and posttests (HVPT group \( n = 12 \), No Training group \( n = 3 \)).

**The discrimination task.** Recall that the first part of the research question was whether HVPT via English Accent Coach leads to gains in perception. The purpose of this first set of descriptive analyses was to determine whether the HVPT group improved their perception of the target vowels following training, and whether these gains can reasonably be attributed to the training. First, d-prime scores were calculated from the proportion correct data for both the pretest and the posttest. D-prime is a measure of sensitivity that factors out individual response bias, such that d-primes of zero indicate chance performance. All presented scores are d-primes. Because the groups had such

<table>
<thead>
<tr>
<th>Task</th>
<th>Pre</th>
<th>Post</th>
<th>Both</th>
<th>Tokens Trained</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXB</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>( m = 2125 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( SD = 489 )</td>
</tr>
<tr>
<td>WL</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>( m = 2067 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( SD = 561 )</td>
</tr>
<tr>
<td>AB</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>( m = 1391 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( SD = 247 )</td>
</tr>
<tr>
<td>WL</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>( m = 1345 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( SD = 264 )</td>
</tr>
<tr>
<td>AXB</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>0 (Control)</td>
</tr>
<tr>
<td>WL</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>0 (Control)</td>
</tr>
</tbody>
</table>
small and unequal numbers of participants with usable AXB data, inferential statistics were not appropriate for comparing the two groups.

To determine whether 1) there were any differences between the pretest and the posttest and 2) whether the differences could be attributed to the training, overall means and standard deviations were calculated for each group at each time. Both groups demonstrated very low detectability of the target contrasts at pretest (HVPT group: \( m = 0.62, SD = 0.69 \); No Training group: \( m = 0.25, SD = 0.82 \)) and at posttest (HVPT group: \( m = 0.84, SD = 0.81 \); No Training group: \( m = 0.71, SD = 0.84 \)). These results, shown in Figure 3.3, do not appear to indicate that the HVPT group improved more than the No Training group.

Because participants in the HVPT group received small and variable amounts of training, next we investigated whether participants who completed more training also demonstrated greater improvement on the discrimination task. Amount of improvement was operationalized as the difference between pretest and posttest \( d^{\prime} \)-primes. Results of a correlation analysis (shown in Figure 3.4) showed that there was a significant correlation between the amount of training and the amount of improvement on the AXB task for the HVPT group \( (r = .532, p = .028) \), suggesting that the training is indeed contributing to participants’ sensitivity to the contrast. The lack of overall improvement in the AXB task may be due to not having had enough training, as participants received considerably less training than what was originally intended (18 sessions totaling 3600 tokens). The very low proficiency of the participants may also have meant that the first few training sessions were used primarily to learn how the program worked. There are a number of other possible contributing explanations for the lack of demonstrated gains. The
Figure 3.3 E1 AXB mean d ‐ prime scores by group and time. Error bars represent 95% confidence interval.

Figure 3.4 Scatterplot of the number of tokens that each E1 HVPT participant was trained on and their improvement on the AXB task with a superimposed line of fit.
experimental setting was unfamiliar for all of the coaches and they had a particularly
difficult time understanding what was asked of them at the pretest. The control class did
not have access to an interpreter at the pretest (but they did at the posttest). As such, it
would not be surprising for there to have been a large task effect; that is, experience with
the task would result in better scores for both groups. Finally, it is important to recognize
that an identification task (such as the one used for the training) and a discrimination task
are different tasks tapping different levels of phonological representation, and
demonstrating gains from pretest to posttest requires generalization of learning from one
task to another. It is worth noting that participants’ accuracy in the training sessions
trended upward over the course of the experiment.

**The word learning task.** The second part of the research question asked whether
participants would be able to exploit any perception gains for lexical acquisition. As with
the discrimination task, small and unequal numbers of participants in the HVPT and No
Training groups renders inferential statistical analyses inappropriate, so only descriptive
analyses will be presented. Near zero d-primes indicate that the difference between
matched and mismatched items were not detectable for either group at pretest (HVPT
group: \( m = -0.35, SD = 0.76 \); Control group: \( m = 0.00, SD = 0.00 \)) or at posttest (HVPT
group: \( m = -0.21, SD = 0.40 \); Control group: \( m = 0.12, SD = 0.20 \)). These results,
summarized in Figure 3.5, do not seem to indicate any improvement in the ability to learn
words containing the target contrasts for either group.
Discussion

As discussed above, this study differed from earlier studies in a number of ways, many of which may have contributed to the apparent lack of improvement following HVPT. With regard to the training, participants completed considerably fewer training sessions than in most earlier studies (see, e.g., Thomson & Derwing, 2016; c.f. Thomson 2012) and they did so within the constraints of an established ESL instructional setting. This meant that it was restricted to the availability of the computer lab (which was not the normal classroom), the ability for them to get logged on to University computers (which, as visiting rather than matriculated students, was restricted), and the confines of both the university course schedule (as mentioned earlier, classes were canceled campus-wide in the middle of the experiment) and the time and day of the class (one of the experimental sessions occurred back-to-back with an off-campus sporting event that the coaches were required to attend, so many coaches were either very late or opted out of returning to campus).

A number of factors may have conspired to result in such low numbers of participants, especially in the control class (in which only three of 25 students
successfully completed the WL pre- and posttests). With regard to the testing, participants completed this within a class session as well, providing a testing environment much different from that of a laboratory. One of the classes was fairly active, resulting in a distracting testing environment. Students were talkative and excitable. In the control class, some students expressed doubt about participating (though they were reminded that it was voluntary and they could leave at any time), and once this opinion was voiced, nearly a dozen students who had been preparing to begin the experiment instead stood up and left. One of the testing sessions occurred on a day when the class was taught by a different instructor in which students appeared to be less concerned about remaining for the entire class period compared to when other instructors were present. For all groups, there seemed to be a crowd mentality, such that as soon as people started finishing others began to rush or decided to abandon the task in order to leave with their friends.

With regard to the participants, these ESL learners had a much lower proficiency in English than those of earlier studies (see, e.g., Thomson & Derwing, 2016), and had been in an English-speaking country for only a few weeks at the beginning of the experiment. Participants were in general older than typical college-age students because they were professionals visiting for a training program. Participants were in the United States for professional skills training rather than interest in learning English; all of them had established careers in China that they would be returning to. Finally, since they were not regular matriculated American university students, they were not used to working with researchers and they were easily frustrated by the tasks. These differences between the present study and earlier studies can help us understand the limitations of HVPT, and they also contribute to our understanding of the challenges that must be navigated when
applying methodologies of laboratory studies to real-world language learning settings.

**Experiment 2**

The second study (E2) is meant to represent a compromise between strictly laboratory experiments and the classroom-based pseudo-experiment 1. Participants completed the training on their own time in a location of their choosing. They were required to complete a specified number of training sessions over the course of 3 weeks. The pre- and posttests were conducted individually in the University of Utah’s Speech Acquisition Lab.

**Participants**

Participants in the second experiment were recruited from ESL and linguistics classes at the University of Utah. Participants’ demographic data are summarized in Table 3.5. Participants were ages 18-35 \((m = 22.8, SD = 3.93)\), had been in the United States for between 2 months and 6 years, and had intermediate English proficiency. They were assigned to one of two conditions: HPVT via English Accent Coach \((n = 10)\) or No Training \((n = 10)\). Nine additional participants were excluded for not returning for the posttest or timing out (not responding within 2500 ms) on more than 50% of one of the test tasks. Eleven of the participants with usable data were given the option to participate in the training and 10 opted to do so; the eleventh is represented in the No Training group. The remaining nine members of the No Training group were not given the option to participate in the training. All participants were compensated for the testing with either course credit or $10/hour of their time. Participants were compensated for the training on
a sliding scale such that as they completed more training, they were compensated at a higher rate ($10 for the first week, $15 for the second, and $20 for the third).

**Training Stimuli and Methods**

Training stimuli and sessions on English Accent Coach were identical to those for Experiment 1. However, training procedures differed in two important ways. First, HVPT group participants were asked to do the training on their own time in a location of their choosing over the course of 3 weeks, rather than in a structured classroom environment. Second, participants were required to complete 18 training sessions totaling 3600 tokens over the time period (approximately six sessions per week) and were told not to complete more than three sessions in 1 day. Participants in the No Training group did not receive any training.
Testing Stimuli and Methods

The pre- and posttests were identical to those in Experiment 1, except that each participant was tested individually, and participants completed the tasks at computers in sound-attenuated booths. No interpreting services were provided.

Analysis and Results

All raw data were converted to d-prime scores, as in the first experiment. The d-primes were first submitted to an ANOVA with group as the between-subjects factor (two levels: HVPT and No Training), task type as a within-subjects factor (two levels: discrimination and word learning), and time as a within-subjects factor (two levels: pretest and posttest). Results are summarized in Table 3.6. There was a significant main effect of task type, $F(1, 18) = 49.36, p < .0005, \eta_p^2 = .733$, no interaction between task type and group, $F(1, 18) = 1.852, p = .190, \eta_p^2 = .093$, no main effect of time, $F(1, 18) = 1.472, p = .241, \eta_p^2 = .076$, no main effect of group, $F(1, 18) = 1.198, p = .288, \eta_p^2 = .062$.

Table 3.6 Summary of results from ANOVA of group x task type x time. Asterisk denotes significant effect.

<table>
<thead>
<tr>
<th>Factor</th>
<th>$F(1,18)$</th>
<th>p</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>1.198</td>
<td>.288</td>
<td>.062</td>
</tr>
<tr>
<td>*Task Type</td>
<td>49.36</td>
<td>&lt;.0005</td>
<td>.733</td>
</tr>
<tr>
<td>Time</td>
<td>1.472</td>
<td>.241</td>
<td>.076</td>
</tr>
<tr>
<td>Group x Task Type</td>
<td>1.852</td>
<td>.190</td>
<td>.093</td>
</tr>
<tr>
<td>Group x Time</td>
<td>1.447</td>
<td>.245</td>
<td>.074</td>
</tr>
<tr>
<td>*Task Type x Time</td>
<td>59.416</td>
<td>&lt;.0005</td>
<td>.767</td>
</tr>
<tr>
<td>Group x Task Type x Time</td>
<td>2.391</td>
<td>.139</td>
<td>.117</td>
</tr>
</tbody>
</table>
.062, no interaction between time and group, $F(1, 18) = 1.447, p = .245, \eta^2_p = .074$, a significant interaction of task type and time, $F(1,18) = 59.416, p < .0005, \eta^2_p = .767$, and no three-way interaction, $F(1, 18) = 2.391, p = .139, \eta^2_p = .117$.

Because there was a significant main effect of task type and an interaction between task type and time, the remaining analyses will be split by task type. As in Experiment 1, the two tasks help us to answer two different parts of the research question: the discrimination task reveals whether the participants’ perception of the target contrasts improved following HVPT under these conditions, and the WL task reveals whether the participants were able to exploit the perception gains following HVPT for word learning. The original hypothesis was that the HVPT group would outperform the No Training group overall, but the lack of a main effect of group or any interactions with group suggest that this is not the case.

**The discrimination task.** The first research question was whether the HVPT group would improve their perception of the target vowels on the discrimination task following training, and whether these gains can be attributed to the training. To determine whether the groups perform differently from each other on the pre vs posttest, the d-prime data were submitted to an ANOVA with group as the between-subjects factor (two levels: HVPT and No Training) and time as the within-subjects factor (two levels: pretest and posttest). The results are summarized in Figure 3.6. There was a significant main effect of time, $F(1,18) = 23.031, p < .0005, \eta^2_p = .561$, in the expected direction, with posttest d-primes higher than at pretest, but no significant main effect of training group, $F(1,18) = 1.971, p = .177, \eta^2_p = .099$, or significant interaction of time of test and training group, $F(1,18) = 3.367, p = .083, \eta^2_p = .158$. 
Because we are interested in whether the HVPT group shows greater improvement on the perception task than the No Training group, we also looked at whether, when accounting for any pretest differences among groups, there was a difference in how the groups performed on the posttest. To do so, we performed an ANCOVA with training group as the independent variable (two levels: HVPT and No Training), posttest d-prime as the dependent measure, and pretest d-prime as a covariate. There was a significant effect of group $F(1,17) = 5.413, p = .033$, $\eta^2_p = .242$, indicating that the trained group was indeed ultimately more accurate at discriminating the contrasts in an AXB task than the untrained group. Descriptive statistics are provided in Table 3.7.

Because we have data that allow us to look at discrimination by vowel contrast, we did a final analysis to see whether HVPT differentially affected the discrimination of the trained vowels. The d-prime data were first submitted to an ANOVA with group as the between-subjects factor (two levels: HVPT and No Training) and time (two levels: pretest and posttest) and vowel contrast (three levels: /ɪ-ɛ/, /ɛ- æ/, and /ɑ-ʌ/) as within-
Table 3.7 E2 AXB mean d-prime scores by time and token type. Standard deviations in parentheses.

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>HVPT</td>
<td>0.88 (0.52)</td>
<td>1.86 (0.69)</td>
</tr>
<tr>
<td>No Training</td>
<td>0.76 (0.89)</td>
<td>1.20 (0.66)</td>
</tr>
</tbody>
</table>

subjects factors. There was a main effect of contrast $F(2, 36) = 13.608, p<.0005, \eta_p^2 = .431$, a main effect of time $F(1,18) = 41.467, p<.0005, \eta_p^2 = .765$, no main effect of group $F(1,18) = 1.221, p = .284, \eta_p^2 = .064$, an interaction between time and group $F(1,18) = 5.678, p = .028, \eta_p^2 = .240$, an interaction between time and contrast $F(2,36) = 8.101, p = .011, \eta_p^2 = .310$, and a three way interaction $F(2,36) = 3.365, p = .046, \eta_p^2 = .157$. Because there was a three-way interaction between group, time, and contrast, the Data were next split by group to see whether the training differentially affected the vowels. The data for each group were submitted to an ANOVA with time (two levels: pretest and posttest) and contrast (three levels: /ɪ-ɛ/, /ɛ- æ/, and /ɑ-ʌ/) as a within-subjects factors. The results are summarized in Figure 3.7 and Table 3.8.\(^5\)

For the HVPT group, as expected, there was a main effect of time ($F(1,9) = 31.952, p<.0005, \eta_p^2 = .780$), a main effect of contrast ($F(2,18) = 5.847, p = .011, \eta_p^2 = .394$), and an interaction between time and contrast ($F(2,18) = 18.189, p<.0005, \eta_p^2 = .669$). Planned comparisons indicate that performance on all three contrasts improved

\(^5\) Descriptively, the participants from E1 show the same patterns of perception among the vowels, with the lowest accuracy of /ɛ-æ/ and similar, higher accuracy for the other two contrasts.
Figure 3.7 E2 AXB mean d-prime scores by contrast and time. Error bars represent 95% confidence interval.

Table 3.8 E2 AXB mean d-prime scores by contrast and time. Standard deviations are in parentheses.

<table>
<thead>
<tr>
<th>Group</th>
<th>Contrast</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/ɪ-ɛ/</td>
<td>-0.29 (0.78)</td>
<td>2.41 (0.87)</td>
</tr>
<tr>
<td>HVPT</td>
<td>/ɛ-æ/</td>
<td>0.40 (0.54)</td>
<td>1.31 (0.81)</td>
</tr>
<tr>
<td></td>
<td>/ɑ-ʌ/</td>
<td>1.03 (0.61)</td>
<td>2.05 (0.84)</td>
</tr>
<tr>
<td>No Training</td>
<td>/ɪ-ɛ/</td>
<td>0.67 (0.51)</td>
<td>1.73 (0.86)</td>
</tr>
<tr>
<td></td>
<td>/ɛ-æ/</td>
<td>0.07 (1.05)</td>
<td>0.49 (0.53)</td>
</tr>
<tr>
<td></td>
<td>/ɑ-ʌ/</td>
<td>0.79 (1.02)</td>
<td>1.74 (1.14)</td>
</tr>
</tbody>
</table>
from pre to posttest (/i-ε/ $F(1,9) = 58.988, p<.0005, \eta^2_p = .868$; /ɛ- æ/ $F(1,9) = 5.128, p = .050, \eta^2_p = .363$; /α-ʌ/ $F(1,9) = 21.622, p = .001, \eta^2_p = .706$). Performance differed for each of the three contrasts at pretest, such that d-prime scores for /i-ε/ were significantly lower than for both /ɛ-æ/ ($F(1,9) = 10.755, p = .010, \eta^2_p = .544$) and /α-ʌ/ ($F(1,9) = 18.153, p = .002, \eta^2_p = .669$); d-prime scores for /ɛ-æ/ were also significantly lower than for /α-ʌ/ ($F(1,9) = 8.892, p = .015, \eta^2_p = .497$). At the posttest, d-prime scores for /i-ε/ were higher than for /ɛ-æ/ ($F(1,9) = 15.455, p = .003, \eta^2_p = .632$) and as good as for /α-ʌ/ ($F(1,9) = 2.251, p = .168, \eta^2_p = .200$); d-prime scores for /ɛ-æ/ remained lower than for /α-ʌ/ ($F(1,9) = 5.515, p = .043, \eta^2_p = .380$). These results suggest that the HVPT was most effective for the /i-ε/ contrast.

For the No Training group, there was likewise a main effect of time ($F(1,9) = 33.218, p<.0005, \eta^2_p = .787$), a main effect of contrast ($F(2,18) = 9.432, p = .002, \eta^2_p = .512$), but no interaction between time and contrast ($F(2,18) = 0.843, p = .447, \eta^2_p = .086$), as expected since these participants did not receive any training. Planned comparisons showed that performance on /i-ε/ improved from pre to posttest ($F(1,9) = 7.977, p = .020, \eta^2_p = .470$) as did performance on /α-ʌ/ ($F(1,9) = 8.371, p = .018, \eta^2_p = .482$); however there was no improvement on /ɛ-æ/ ($F(1,9) = 1.97, p = .194, \eta^2_p = .180$). At pretest there was a difference between performance on /ɛ-æ/ and /α-ʌ/ ($F(1,9) = 7.382, p = .024, \eta^2_p = .451$), but not between the other two pairs of contrasts ($F(1,9) = 1.789, p = .214, \eta^2_p = .166$; $F(1,9) = 0.073, p = .794, \eta^2_p = .008$). At posttest, performance on /ɛ-æ/ was significantly lower than for both /i-ε/ ($F(1,9) = 28.815, p<.0005, \eta^2_p = .762$) and /α-ʌ/ ($F(1,9) = 12.395, p = .007, \eta^2_p = .579$), but there was no difference between d-prime scores for /i-ε/ and /α-ʌ/ ($F(1,9) = 0.001, p = .978$,
The word learning task. The second research question was whether learners would be able to exploit the perceptual gains demonstrated in the discrimination task for learning new words. As with the discrimination task, the word learning d-prime data were submitted to an ANOVA with group as the between-subjects factor (two levels: HVPT and No Training) and time as the within-subjects factor (two levels: pretest and posttest). Results are summarized in Figure 3.8. There was no significant main effect of time ($F(1,18) = 4.033, p = .060$), no significant main effect of group ($F(1,18) = .631, p = .437$), or a significant interaction of time and group ($F(1,18) = .665, p = .426$). This suggests that perceptual gains from HVPT, at least of this type and over 3 weeks, may not be able to be used in a higher-level task such as word learning. However, it is worth noting that the HVPT stimuli were only CV syllables and the WL stimuli were CVC syllables. As such, improvement on the WL task required twofold generalization from the training: between both token types and task types. A WL task using trained stimuli may produce different results.

*Figure 3.8* E2 WL mean d-prime scores by group and time. Error bars represent 95% confidence interval.
Discussion

The results of Experiment 2 indicate that HVPT training via English Accent Coach done under these particular conditions can result in improved perception of difficult L2 contrasts in as few as 3 weeks, suggesting that this type of training may be an accessible and effective resource for language learners who desire extra practice outside of their regular language classes.

Under these training conditions and time frame, it does not appear that participants are able to extend these perception gains to a higher level task such as learning new words containing the difficult vowels. It may be that there was not enough training or that more time is required for the new representations to be robust to lexical encoding. Alternatively, it is possible that the establishment and use of new category representations are not directly connected, and the ability to exploit phonetic knowledge for word learning may require something more than perceptual training. This word learning task differed from many previous word learning experiments in that the artificial lexicon in the first phase did not contain minimal pairs. In this way, although they were told in the instructions to pay attention to the vowels, it was not necessarily obvious that they would need to encode fine-grained phonetic detail as they were learning the words in order to be successful in the third phase. Perhaps if the artificial lexicon had included minimal pairs that they may have been more successful. The decision not to include minimal pairs was twofold: it is more ecologically valid (in that people do not necessarily learn vocabulary in minimal pairs) and it meant that the pretest and posttest were not identical but the stimuli were maximally similar.
General Discussion

This study investigated whether real ESL learners were able to: (a) improve their perception of difficult L2 vowel contrasts via 3 weeks of an online high variability training procedure, and (b) exploit any posttraining gains in perception to learn new words containing the trained vowels. The same training procedure was implemented with two different populations and in two different settings: (1) with low-proficiency Chinese professionals (university sports coaches) in an ESL class setting and (2) with moderate-proficiency native Mandarin-speaking college students on their own time. These two situations were meant to represent implementation of HVPT as either part of a course or extracurricular practice; both were chosen to extend earlier laboratory research on HVPT to more ecologically valid settings.

The coaches in the ESL class setting demonstrated a correlation between the amount of training they received and discrimination improvement from pretest to posttest, indicating that HVPT contributed to their perception of the difficult L2 contrasts. Despite this, they did not demonstrate improvement in either the discrimination task or the word learning task. The learners who did the training on their own time ended up outperforming a control group on the discrimination task, suggesting that an online implementation of HVPT can improve the perception of difficult contrasts for ESL learners in as few as 3 weeks. They also did not demonstrate improvement on the word learning task, which may have been a result of the twofold generalization required (both between token types and task types). It’s worth noting that training the vowels in all possible CV contexts may have provided too much variability, and that these learners—particularly those in E1 who received less training—didn’t receive enough exposure to
the contexts used in the test tasks to be able to show the expected gains. In other words, if training is limited to only the CV contexts tested, therefore providing more exposure to the tested contexts, learners may demonstrate larger gains. The disconnect between discrimination improvement and WL lack of improvement may suggest that low-level perception abilities and the ability to use this perceptual knowledge to learn new words may not develop in parallel or over such a short period of time. That is, successfully learning words containing difficult-to-perceive phonemes may take more time, more training, or a different kind of exposure.

A major area of interest in the HVPT literature is generalization from the particular features of the training stimuli to new test stimuli. Much research has demonstrated improved perception following HVPT that generalizes to new talkers but not new words (see, e.g., Qian, Chukharev-Hudilainen, & Levis, 2018). In our study, training only included CV syllables, while the discrimination task included both CV and CVC syllables. It is possible that it was an inability to generalize learning from CV to CVC syllables affected the overall improvement (or lack thereof for the participants of E1) on the discrimination task. To investigate this, we performed an ANOVA for each group with token type (two levels: CV and CVC) and time (two levels: pretest and posttest) as independent variables. In E1, there was no difference in either group for token type (HVPT: $F(1, 16) = .461, p = .507, \eta_p^2 = .028$; No Training: $F(1, 6) = .531, p = .494, \eta_p^2 = .081$) or interaction between time and token type (HVPT: $F(1, 16) = .054, p = .819, \eta_p^2 = .003$; No Training: $F(1, 6) = .014, p = .910, \eta_p^2 = .002$). In E2, there was no main effect of token type $F(1, 9) = .080, p = .783, \eta_p^2 = .009$ or interaction between token type and time $F(1,9) = 1.376, p = .271, \eta_p^2 = .133$ for the HVPT group either.
Unexpectedly, the No Training group data showed a main effect of token type
\[ F(1,9) = 13.212, \ p = .005, \ \eta^2_p = .595 \] but no interaction between token type and time
\[ F(1,9) = .339, \ p = .575, \ \eta^2_p = .036, \] indicating that there was a systematic difference between perception of CV and CVC tokens for the No Training group that was maintained between the pretest and the posttest. Perhaps even more surprisingly, they showed an advantage for the CVC syllables. This may have been a result of the fact that many of the vowels in this study are lax vowels, which do not occur in open syllables in English, although it is worth noting that only one of the four groups showed this advantage (though descriptively it is the overall pattern for all groups). Descriptive statistics are summarized in Table 3.9. Despite this unexplained finding, the overall results show that HVPT using only CV syllables did not differentially affect performance on CV and CVC syllables at test. The lack of interaction between token type and time in this study contributes to the question of the extent to which perception training can generalize to new contexts; adding a final consonant to a CV syllable does not appear to cause difficulty (this despite Thomson, 2005, finding that presence of a coda consonant

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Group</th>
<th>Pretest</th>
<th>CVC</th>
<th>Posttest</th>
<th>CVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>HVPT</td>
<td>0.624</td>
<td>0.666</td>
<td>0.823</td>
<td>0.914</td>
</tr>
<tr>
<td></td>
<td>(0.729)</td>
<td>(0.747)</td>
<td>(0.862)</td>
<td>(0.976)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No Training</td>
<td>0.186</td>
<td>0.367</td>
<td>0.486</td>
<td>0.645</td>
</tr>
<tr>
<td></td>
<td>(0.921)</td>
<td>(0.761)</td>
<td>(1.041)</td>
<td>(0.608)</td>
<td></td>
</tr>
<tr>
<td>E2</td>
<td>HVPT</td>
<td>0.944</td>
<td>0.829</td>
<td>1.919</td>
<td>2.120</td>
</tr>
<tr>
<td></td>
<td>(0.624)</td>
<td>(0.515)</td>
<td>(0.767)</td>
<td>(0.945)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No Training</td>
<td>0.577</td>
<td>1.597</td>
<td>1.462</td>
<td>2.233</td>
</tr>
<tr>
<td></td>
<td>(0.982)</td>
<td>(0.762)</td>
<td>(0.934)</td>
<td>(0.669)</td>
<td></td>
</tr>
</tbody>
</table>

*Table 3.9. Mean AXB d-prime scores for all groups by token type and time. Standard deviations are in parentheses.*
caused difficulty in Mandarin speakers for perceiving English vowels), even for the very low proficiency coaches.

Another topic of interest in L2 perception literature is the variable difficulty of learning different L2 sounds (i.e., that not all sounds are equally easy to learn). Thomson, Nearey, and Derwing (2009) investigated the Mandarin and English vowel inventories for phonetic similarity, and found that English /æ, ɑ, ʌ/ all overlap with a Mandarin /a/ category, while /ɛ/ was almost completely distinct from any Mandarin category. Interestingly, /ɪ/ was most phonetically similar to a Mandarin /ɛ/ category. Based on this analysis, the Speech Learning Model (Flege, 1995) would predict that the /ɑ-ʌ/ contrast would be most difficult for Mandarin speakers to discriminate. However, overall the learners in this study had most difficulty discriminating /ɛ-æ/.

One possible explanation for this discrepancy is that while learners were trained using Canadian English talkers, they were tested using Utah English talkers. It is possible that the productions of these vowels are sufficiently different between these populations that learners were not able to generalize from one set of voices to the other. In order to determine the similarity of the vowels across these populations, five tokens from each of the vowels were randomly selected from each of the three test voices and from three randomly selected voices from English Accent Coach. For each token, the F1 and F2 of the steady state of the vowel were measured. Figure 3.9 shows the vowel spaces of the respective groups. While the training and test vowels were very similar for /ɛ, æ, ɑ/, the training and test vowels for /ɪ/ and /ʌ/ did diverge some. These differences may have meant that the testing of these two vowels was not done with sufficiently comparable stimuli to determine the effectiveness of training.
Any differences in testing and training voices could not be the whole story, however, since the HVPT and No Training groups did not show clear differences in the relative pretest accuracy or pre- to posttest improvement of the contrasts. Thomson et al. (2009) also showed that /æ/ is much less acoustically similar to the Mandarin /a/ vowel than English /a/ and /ʌ/, and may therefore be considered an “uncategorized” vowel. In that case, both /æ/ and /ɛ/ occur largely in an area where there is no Mandarin equivalent, and the uncategorized-uncategorized pair was more difficult than two vowels that assimilate to a single category with differential category goodness. It is worth noting that despite these difficulties, learners were able to improve on all three contrasts over the course of the experiment.

Together, Experiments 1 and 2 demonstrate that HVPT via an online implementation may not necessarily be robust to the chaos of a truly ecologically valid
setting. That is, in order for learners to demonstrate significant improvement in the
discrimination of vowels over the course of a relatively short time period, some protocols
are required. However, in both experiments learners demonstrated that training did have
an effect; amount of training was correlated with improvement in E1 and the training
group in E2 ended up outperforming the control group.

This study used an AXB discrimination task to test perception, and the results
provide evidence that training via an identification task is able to generalize to another
perception task. To determine the utility of HVPT, future research should continue to
investigate to what extent learning can generalize to other linguistic tasks, including
communicative tasks, and the time course over which perception can be employed for
higher level tasks.

This study also revealed a disconnect in that learners demonstrated improvement
on the perception task that was not able to generalize to a lexical task, despite the
similarity of the two tasks (but see Pater, 2003, for an argument that the tasks may not
have been closely enough related) and the fact that aural stimuli were exactly the same in
the two tasks. The discrepancy in results between the two task types suggest that low-
level perception and the use of perceptual knowledge for higher level tasks may not
develop simultaneously, but this study doesn’t disambiguate between the possible causes
of this (e.g., encoding phonological information in the lexicon may require more time or
other kinds of exposure). A longer period of training or a greater number of training
sessions over a similar period could help determine whether participants in this study
simply had not yet reached the “saturation” necessary for applying the perception
knowledge to a higher-level task. Nishi and Kewley-Port (2007) suggest that training
using a full vowel set may result in more robust generalization than training on a subset of difficult vowels, so full vowel set training may also better facilitate lexical encoding of the difficult contrasts.

Finally, this study was an initial attempt to respond to the call for research investigating how laboratory research can be used in real-world language learning settings. HVPT has been identified as an area that has pedagogical potential (Levis, 2016), and a freely available online program such as English Accent Coach is an accessible option for learners. This study sought to investigate the potential for English Accent Coach’s use both in the classroom and as extra practice. The cause for theoretical researchers’ resistance to classroom-based research quickly became clear. First, it requires the involvement of a lot more people: program directors, instructors, and university room schedulers all needed to be on board in addition to the learners themselves in order to make the study happen. Rooms full of participants resulted in reduced participation due to distractions and group mentality. Time was limited by the university’s schedule, students’ attendance, and other extraneous factors. It also made it difficult to ensure that all participants clearly understood what they were being asked to do, as one-on-one experimenter-participant interaction was limited.

References


Transition from Speech Sounds to Spoken Words, 167(224), 233-277.


CHAPTER 4

THE STATE OF THE FIELD AND FUTURE DIRECTIONS

Introduction

Because phonology is language-specific, it must be learned. As such, the input that a learner receives must provide access to the required information for phonological category formation. There are a number of different features of the input that have been identified as potentially useful for the acquisition of both first and second language phonology. This dissertation focused on two of these potential contributors: distributional information and variability.

Distribution-Based Learning and L2 Allophonic Acquisition

Distributional learning mechanisms have been proposed for both native and second language phonological acquisition. These proposals posit that learners (infants and adults) track distributional information about phones in their input, including how often they are produced, where they occur in a phonetic space, and the phonological environments in which they appear. Infant exploitation of distributional information has been demonstrated for the acquisition of phonemes (e.g., Maye, Werker, & Gerken, 2002) and for grouping allomorphic alternants into the same category (White, Peperkamp, Kirk, & Morgan, 2008). Some research has suggested that adults are also
able to take advantage of distributional information for phoneme acquisition (e.g., Hayes-Harb, 2007; Maye, 2000).

Peperkamp, Pettinato, and Dupoux (2003) hypothesized that adults might be able to use both statistical (monomodal or bimodal) and phonological (overlapping or complementary) distributional information together to infer whether nonnative phones belong to one or two categories. The results showed that there was a significant difference between pre-post performance for the group that received bimodal statistical information about phones in overlapping distribution (i.e., the exposure that was meant to simulate phones belonging to two separate categories), although there was not a significant interaction between exposure type and the pre-post difference in performance on a discrimination task. Shea and Curtin (2010) investigated whether, as learners gain experience with the language, they demonstrate knowledge of the relationship between L2 allophones and their conditioning phonological environments. The results showed that high-proficiency learners and native speakers associated the correct alternant with stress, while low-proficiency learners and participants naïve to the language did not show this preference. The authors attributed this differential sensitivity to a phonological distributional mechanism, however because the learning of the alternation occurred outside of the experimental setting, it is unclear which features of the input were used to make this association.

The Chapter 2 study sought to isolate phonological distributional information to determine whether it alone could explain adult acquisition of second language allophones. Relatively inexperienced listeners were exposed to a novel L2 contrast in either overlapping or complementary distribution. We hypothesized that if adults are able
to use this information to infer whether the two phones were related as allophones of a single phonemic category or belonged to two different categories, the participants who were exposed to the contrast in overlapping distribution would perform more accurately on an ABX task than those who were exposed to the contrast in complementary distribution, in accordance with the empirical observation that native speakers are less sensitive to the distinction between allophones than phonemes (Boomershine, Hall, Hume, & Johnson, 2008). The results did not show any difference between the two exposure conditions with regard to the discrimination performance on the target contrast, which was consistent with the null results of Peperkamp et al. (2003), suggesting that phonological distributional information alone may not be sufficient for adults to infer the phonological relationship between a novel pair of phones.

As the literature on adult allophonic acquisition is relatively sparse, the primary contributions of this study were methodological with regard to (1) the stimuli used in the training and (2) the dependent measure. (1) Unlike in Peperkamp et al. (2003), the stimuli were all naturally produced rather than spliced together or artificially manipulated into a continuum. This was meant to be more ecologically valid, reducing unnatural cues and providing input more similar to that of real-world language learning. There were limitations to this decision, however, such as the difficulty of producing allophones in illegal contexts. Because we wanted to completely isolate the variable of phonological distributional information, we needed to ensure that productions in illegal contexts weren’t systematically different from those in legal contexts via a phonetic analysis. However, research has indicated that when two phones are allophones of the same phonetic category, they differ not only in perception but also in production as compared
to when two phones belong to distinct categories. As such, it is possible that we exaggerated the statistical distribution of the target phones compared to their distribution in a natural language. (2) An ABX task was chosen as the perceptual measure because it is traditionally understood as relying on a higher level of phonological representation than an AX task (as was used in Peperkamp et al., 2003). However, because all of our tokens were produced by a single speaker, participants may have been able to use an acoustic strategy in which they ignored the first token and compared only the last two tokens, effectively performing an AX task anyway. The ABX task was also a less metalinguistic task—comparing two aural stimuli—than the task used in Shea and Curtin (2010)—deciding which syllable is stressed. As such, we believed that this type of task would tap into the level of representation necessary to demonstrate knowledge of allophony.

The present study represents an initial attempt to investigate whether adults can use phonological distributional information in naturally produced stimuli to learn that two phones are related as allophones of a single phoneme. It did not support the hypothesis that phonological distributional information alone is sufficient for allophone acquisition. Many questions remain, and there is much to build upon, as discussed next.

We are unfortunately unable to disambiguate among the possible causes for a null effect; the lack of an effect in our study does not tell us that learners are unable to use phonological distributional information for allophonic acquisition but only that the conditions we provided did not facilitate the exploitation of phonological distributional information in the exposure for AXB performance. The question remains as to whether adult L2 learners are able to use phonological distributional information in the input for
allophonic development. The reasons to continue investigating this as a possible mechanism are compelling: infant research suggests that such a mechanism is used in first language development, and similar mechanisms have been empirically supported as maintaining efficacy for adult L2 phonemic development (see, e.g., Hayes-Harb, 2007). It is possible that adults require more evidence (i.e., more instances) of phonological distribution for allophones than statistical distribution for phonemes, and therefore simply needed more exposure than was provided in this study. Seidl and Cristia (2012) argue that infants’ development of allophones involves both complementary distribution and phonetic similarity. For adults with an established (native) phonological system, ‘phonetic similarity’ may be calculated somewhat differently, taking their phonology into account. For example, although [b] and [β] are phonetically similar enough for allophony (in, e.g., Spanish), English speakers often identified [β] as belonging to a /v/ category (which doesn’t exist in Spanish) in Chapter 2. This two-category assimilation (Best, 1995) would require learners to discover that [β] is not an instance of the native /v/ category but instead is related to /b/ and may result in the calculation for ‘phonetic similarity’ being more complicated for adults than for infants, and therefore may require more exposure to overcome.

It could also be that phonological distributional information alone is not enough, and that it must to be presented with other kinds of information to be useful. Perhaps, as Peperkamp et al. (2003) hypothesized, it is the combination of both phonological and statistical distributional information that clues learners in to the phonological relationships among phones. Alternatively, Peperkamp and Dupoux (2002) argue that the ability to use a distributional mechanism for phonological development would help with
lexical acquisition, but perhaps the reverse is true as well. A third possibility could be that adult learners use their lexical knowledge to make phonological inferences. It would be useful to better understand the time course of L2 allophonic acquisition in order to understand how the phonology and the lexicon might influence each other. An area of great interest within L2 phonological acquisition research is the role of orthographic information, one area of input that is unique to the L2. An interesting question would be whether phonological distribution together with written input indicating that multiple phones are represented by the same orthographic symbol can help participants map alternants to a single phoneme.

An important question for moving this area of study forward relates to the extent to which allophones actually occur in the distributions as defined in theoretical phonology. Shea (2014) notes that “many allophonic relationships that were previously characterized as involving complementary distribution are better conceived of existing on a continuum, with binary distribution as a tendency, rather than absolute.” (p. 1338). It’s also worth noting that the idea of learning an L2 allophonic relationship actually covers a number of different relationships that need to be learned: phones from two different L1 categories could be allophonically related in the L2, a phone that doesn’t exist in the L1 could be an allophone of a category to be learned, a phone from an L1 category could be an allophone of a different L1 category in the L2. Each of these alternation types may require different information to be learned, and it is worth studying learners from a variety of L1s and L2s to understand the differences among learning each of these patterns. Are some of these patterns easier or more difficult to learn?

For us to understand the ways first and second language phonological acquisition
are similar or different, we must continue to explore the extent to which the mechanisms that appear to be utilized in infant acquisition can also be accessed for L2 acquisition. The areas of phonological distributional information and allophone acquisition are not yet well understood, and are rife with opportunity for exploration.

**High Variability Phonetic Training and L2 Phonological Acquisition**

High variability phonetic training (HVPT) is a training protocol originally conceived by Logan, Lively, and Pisoni (1991) to test the hypothesis that the variability inherent in speech sounds produced by a variety of talkers and in a variety of phonetic environments may be beneficial for developing robust second language (L2) phonetic categories. It is presumed that this kind of exposure helps learners to sort out which dimensions of variability are relevant and which are irrelevant for distinguishing the difficult L2 sounds in a target-like way. Traditional HVPT involves phonetic identification training, in which participants hear a sound and select what they heard from a number of choices (e.g., images, such as nautical flags [Thomson, 2011] or symbols such as the International Phonetic Alphabet in a vowel chart [Thomson & Derwing, 2016]). After their selection, they receive immediate feedback indicating whether or not their choice was correct.

The first few studies (Logan et al., 1991; Lively, Logan, & Pisoni, 1993) investigated whether this type of training could result in better identification of the target phones both as produced by the trained talkers and contexts and also as produced by new talkers and in new contexts (generalization), and particularly whether HVPT was superior to single-talker training with regard to both identification improvement and
generalization capabilities. Bradlow, Pisoni, Akahane-Yamada, and Tohkura (1997) were the first to investigate whether the identification training could generalize to other kinds of tasks (in their case production). More recently, Ingvalson, Barr, and Wong (2013) provided some evidence that HVPT on lexical tone (suprasegmental phonology) can result in improved ability to learn words containing lexical tone.

Over the last few decades, HVPT has become an area of inquiry in its own right in which speech researchers are interested in determining the scope of utility of this protocol for language learning. HVPT has increasingly been identified as an area of laboratory research that has great potential for use by L2 learners and teachers in the real world, particularly in the area of pronunciation (Levis, 2016).

The studies in Chapter 3 sought to extend the HVPT laboratory research findings to less controlled learner-centered settings by involving real L2 English learners both (a) in an established ESL class and (b) who completed the training in the time and location of their choosing, simulating extracurricular practice. These studies also investigated whether the perception gains from HVPT on segmental phonology could be generalized to a higher level, lexical task. The hypothesis was that learners receiving HVPT would perform more accurately on a task in which they had to learn new words containing the trained difficult L2 phonemes (in our case, vowels) than a control group of learners who did not receive the training. Experiment 1, involving low-proficiency learners who were visiting students in an established ESL class, resulted in participants from both the HVPT and No Training groups showing no improvement on either the perception task or the lexical task. However, there was a correlation between the amount of training that the participants in the HVPT group completed and their improvement on the perception task.
Experiment 2, involving moderate proficiency matriculated students who did the training outside of the laboratory or a class, resulted in the HVPT group outperforming the No Training group on the posttest, though neither group demonstrated improvement on the lexical task. Despite the lack of word learning improvement demonstrated by the language learners in these experiments, there is much evidence to suggest that perception is an important precursor to other linguistic abilities. As such, these experiments represent an essential contribution to the field of L2 phonology as an incremental step toward better understanding of the possible contributions of perceptual learning to higher-level abilities such as word learning.

Generalization

A major area of interest within the HVPT literature is generalization—that is, the extent to which training on one thing can result in improvement on something else. The earliest studies (e.g., Logan et al., 1991; Lively et al., 1993) were interested primarily in generalization to new talkers and new words. While these were not the primary research questions of our study, the design of our study included both of these variables. With regard to talkers, the 30 training voices were all speakers of western Canadian English, while the three testing voices were produced by people who were not only not included in the training voices, but were also speakers of a different variety of English, specifically western United States (Utah) English. An acoustic analysis suggested that the two varieties overlapped substantially for most of the target vowels, though not equally so for all five. Despite these differences, the participants did seem to be able to generalize their learning from the training to the perception test task. Because learners of a second
language living in an immersion setting will likely come into contact with native speakers of multiple varieties, future research may consider the generalization of learning across dialects, particularly with regard to vowels.

Regarding generalization to new words, the HVPT training stimuli in our study contained only CV syllables, but in the perception task, stimuli included both CV and CVC syllables. Neither of the HVPT groups demonstrated an advantage for CV syllables over CVC syllables on the discrimination task, indicating that the participants were able to generalize their perception gains on CV syllables to CVC. Many HVPT studies have shown limited generalizability to new phonetic contexts (see e.g., Qian, Chukharev-Hudilainen, & Levis, 2018; Thomson, 2011; Thomson, 2016), so research directly addressing which types of new phonetic contexts do and do not cause difficulty is called for.

More recently, speech researchers have become interested in the ability of learners to generalize their perception of a subset of phonemes to other, related phonemes. For example, Nishi and Kewley-Port (2007) investigated whether HVPT on a subset of particularly difficult vowels would result in improved perception of all vowels. They found that this was not the case; additionally, participants whose HVPT contained the entire vowel set demonstrated greater gains overall on the full set of vowels and were better able to generalize the learning to new talkers and new words than the group who were trained on the subset. In this study, we only tested learners on the subset of vowels that they learned, and it is an open question as to whether the learners would have also demonstrated improvement on other vowels. There has been relatively little additional research into this type of generalization, and given that it could provide a more efficient
way of learning to perceive difficult sounds, further study may be warranted.

Our primary contribution in the area of generalization was regarding the generalization of the identification training task to other kinds of tasks. First, our test of perception was a discrimination task. Participants who received HVPT did demonstrate that they were able to generalize one perception ability (identification) to another (discrimination), in that the E1 participants showed a correlation between the amount of completed training and improvement on the AXB task, and the E2 HVPT group ended up outperforming the No Training group on the posttest.

Second, our research question directly addressed the generalization of perception gains to a higher-level task. The results showed that none of the four groups of participants (E1 HVPT and No Training, E2 HVPT and No Training) demonstrated any apparent improvement on the lexical task, or indeed any sensitivity at all to the difficult vowels in the words (all d-primes were around zero). The lexical task differed from many earlier artificial-lexicon studies (including that used in Ingvalson et al., 2013) in that it did not teach minimal pairs. This meant that in the word learning phase, participants were taught, for example, [ktʃ] and then on the final test were asked whether the picture they had been taught as [ktʃ] was [kɛʃ], though they had not learned a meaning for [kɛʃ]. It is possible that in previous artificial lexicon studies, it is the presence of minimal pairs that cues participants to recognize that they need to encode fine-grained phonetic detail about difficult phones as they are learning the words in order to be successful in the third phase. We opted to not include minimal pairs as the primary goal of this study was to determine the efficacy of HVPT under less-tightly controlled conditions, and learners outside of a laboratory often do not learn vocabulary in minimal pairs. We may not have provided our
participants with the best-case scenario to demonstrate gains in this area, though, and future research may consider whether teaching minimal pairs might have done so.

On the other hand, the lack of learning in the lexical task may provide preliminary evidence suggesting that perception and the ability to use this knowledge for a lexical task do not develop in parallel. That is, though participants demonstrated perception gains, they did not seem to show equivalent gains in their ability to learn new words. This is an area requiring more investigation, because the implications are important for our understanding of the relationship between L2 phoneme category formation and lexical representations. Darcy et al. (2012) similarly found that category formation and the ability to use phonemic information in lexical development may not develop in parallel, though their results suggested that the order was reverse: that the lexical development may precede stable phoneme category representations. Their participants completed an AXB discrimination task and then a lexical task, though the task they used to tap lexical representations was different from ours; they used a lexical decision task with repetition priming. This may be more similar to the AXB task than our artificial-lexicon task was, and together these tasks could help us understand the hierarchy of lower- to higher-level tasks. Future research might consider comparing the results of these two lexical tasks or investigating if learning from HVPT is able to generalize to a different kind of lexical task, such as lexical decision task. (Note, however, that we could not have used a lexical decision task in E1 because the proficiency of the coaches was too low for this kind of task.)

In order to determine the scope of HVPT’s utility for L2 learners, future research should continue to investigate the extent to which real learners’ perception gains from
HVPT can generalize to other linguistic tasks, such as word learning, segmentation of the speech stream, production, and other more communicative tasks. Results from this study suggest that perception and lexical encoding may not happen in parallel; there is an open question of the time course over which perception can be employed for higher level tasks.

The Intersection of Research and Pedagogy

Another contribution of Chapter 3 is that it is an attempt to extend the robust laboratory findings that HVPT can mediate cross-language speech perception to real L2 learners and the settings in which they commonly find themselves (i.e., in a classroom or looking for extra practice online). As mentioned above, there has been increasing interest in research that bridges the gap between the field of SLA’s theoretical and laboratory understanding of how language is learned and language learning and teaching in the real world. In particular, HVPT has been identified as a good candidate for computer-assisted pronunciation training and for providing practice that complements that which is already provided in regular classroom instruction (Thomson, 2011). Thus, the Chapter 3 studies employed a freely-available online HVPT platform for learning English vowels and consonants, English Accent Coach (EAC; Thomson, 2012).

Experiment 2 demonstrated that learners need to follow only minimal protocol (in this case, completing six 200-token sessions per week, no more than three in a single day) to demonstrate improvement in perception in just 3 weeks. The reduced control of timing and location, which arguably could have resulted in problems (e.g., participants doing all the sessions in the last 2 days of every week, having other noises going on in the background, not using high-quality headphones, multitasking while doing the training),
did not appear to impede learning (at least with regard to perception; it is at least possible that this could have contributed to the lack of effect for word learning). This is an important finding: that HVPT via EAC appears to be amenable to less-controlled situations that simulate something like homework or extracurricular practice. Future research may be interested in the minimum amount of training necessary for gains and for how long continued training results in continued improvement. It is possible that improvement on the lexical task could simply require more training than we required, so it would be reasonable to investigate if with more training, learners are able to show gains on a lexical task (either one like ours or like that in Darcy et al., 2012).

Experiment 1 contributes to our understanding of some of the additional challenges that come with adapting laboratory research to classroom research (and particularly that which occurs over a number of weeks). Conducting human subjects research always involves informed consent. When such research is conducted in a classroom setting, the researcher and teacher must negotiate a number of issues. For example, while our goal is to have as many participants as possible, we need to ensure that participants are not coerced and that they feel like participating is a good use of their time. In the case of the present research, data collection involved the cooperation of six different instructors, as well as a translator to make certain that participants understood what they were agreeing to, and a lot of volunteers to help run the experiment. Their normal classroom was not a computer lab, so training and testing required additional scheduling of rooms, which meant that when participants were unable to complete the goal number of training within the given time, it was not possible to simply push back the posttest date. Unforeseen changes to the university course schedule (in the middle of the
3-week training period, a campus tragedy shut down the University for a day; there could just as easily have been a weather incident) and any number of personal reasons can keep students from class, and therefore from fully participating. Giving instructions to a large group, especially when there are a lot of questions, can take much longer than giving instructions to an individual, and this meant that many willing participants did not finish the pretest, which made their data unusable.

Additionally, the testing environment of a class session is quite a bit different from that of a laboratory. The classes were loud and restless, with lots of movement and chatting throughout the experimental session. Many students appeared to be influenced by their classmates’ classmates, either getting up to leave after one particularly loud student said they didn’t want to participate (though until that point they had been preparing to begin the test), or rushing or giving up altogether after the first few participants finished the tasks. These scenarios are all much less likely in a lab setting, where participants self-select to show up and are usually tested individually. It is possible that some of these challenges could have been mitigated if students had been better prepared for what was going to happen before the first session. Future research might consider having a classroom teacher as a co-experimenter, who understands how research works and the benefits of it and is able to prepare the students for what an experimental session might look like. Future research may also use matriculated ESL students, who are more likely to be familiar with research and whose motivation for improvement is more extrinsic (i.e., the need to be successful in classes for which the primary language of instruction and work is English), so they may feel like participation is a better use of their time than the visiting students in E1 did.
The lack of improvement for both groups in E1 does not mean that HVPT via EAC cannot be useful in a classroom setting. Rather, it is a function of a unique combination of factors—the low-proficiency visiting students, the instructors not having buy-in for the experiment, and a number of other specific details about when and how the class met—that might serve as a useful guide for someone else to know what pitfalls to avoid and the wiggle-room that ought to be built in for this kind of research. This area is wide open, both for better understanding how to apply laboratory research to applied research, and for understanding of the scope of the utility of HVPT for L2 learning and teaching.

References


Thomson, R. I. (2012). English Accent Coach (Version 2.3) [Computer program].
