LEARNING NOVEL VOWEL CONTRASTS: EXPERIMENTAL METHODS IN
CLASSROOM APPLICATIONS


BY

LISA RENEE PIERCE


DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014


Urbana, Illinois


Doctoral Committee:

      Professor Jennifer S. Cole, Chair
      Professor Jose I. Hualde
      Professor Tania R. Ionin
      Professor Chilin Shih

# Abstract

This dissertation reports on perceptual training of tense-lax vowel contrasts in the context of an advanced-level ESL pronunciation class for Chinese and Korean international graduate students. The vowel contrasts were trained under four training paradigms designed to examine the effects of variation due to multiple speakers, different speech rates and coda consonant. Training material consisted of nonce word minimal pairs used to mitigate task complexity related to lexical access and to circumvent the effects of frequency and top-down processing.

Participants completed pre- and post- tests on discriminating vowels in real word minimal pairs and nonce word minimal pairs. Vowel perceptual training took place over the course of six days wherein each day's training consisted of 100 exemplars that students played as much or as little as they wanted; the number of sound files they played was tracked and recorded. Participants finished training at their own discretion and then were tested on a 25-member subset of that day's training tokens.

Chinese and Korean learners trained under one of four training paradigms; Training Paradigm A, multiple speakers, three speech rates; B, multiple speakers, one speech rate; C one speaker, three speech rates; D, one speaker, one speech rate. Chinese and Korean participants who trained in Paradigm A and B (but not C and D) made significant gains in their ability to identify vowel contrasts in real-word minimal pairs. Mandarin speaking participants made significant gains in the nonce-word post-test regardless of training paradigm; Korean participants made significant gains in nonce words in Paradigms A, B, and C but not D. Analysis of the test items indicated that Mandarin speakers made significant gains for tense vowels in training Paradigm A; lax vowels made significant gains in Paradigms B and D. Korean participants who

trained in Paradigm C, saw significant gains for [ɛ], [i], [ɪ], [u], and [ʊ]. They also made significant gains identifying [ɛ] in Paradigm A, but no other training paradigm produced a significant difference between pre- and post-test. In sum, individual vowels were differentially difficult to master depending on the training paradigm, L1 and coda condition.

These findings are considered in terms of the effects of variation in the training material, the influence of the L1 vowel inventory and language specific patterns of co-articulation related coda effects.


Key Words: vowels, perceptual training, nonce words, coda context, SLA, classroom-based research, CALL

## Acknowledgements

I wish to thank my adviser, Jennifer Cole, for her flexibility and patience. I thank my committee members, Jose Hualde, Tania Ionin, and Chilin Shih, not only for their guidance and support, but also for their willingness to be caught up in my enthusiasm for the various side projects I have wanted to explore along the way.

I wish to extend a special thanks to Suyoun Yoon, Tim Mahrt, and Liam Moran who helped with various instantiations of the web-based training application. The final version of the application is credited to Liam, who also trouble shot, tweaked, debugged, and re-coded as needed. Thanks to all the linguistics graduate students who were "the voices".

Finally, I wish to thank: Nagi Khadr for the space and the peace; Mary Ellen Fryer for her efforts and patience; Nesrin Bakir for insisting on transitions and details; Emily Schoenrock for the shiny rocks; Martin Srajeck for his availability and care; Liz Howard for checking up on me; Kopi for letting me be; Dennis Da Silva for understanding; Jennifer Puitt for her enthusiasm; Irene Korn for her willingness to help; Muki, Julio and all the koi; Cindy Overrocker for her care; James Griffith who started everything; HP and LotRs for ongoing support.

**Table of Contents**

# Chapter 1. Introduction

Carroll (2001): *There is, however, not a single shred of evidence that learners can think about the sounds of the speech signal in terms of the constituents and properties which must be acquired if they are to master the phonology of the L2. ... no one says: "I must learn to hear the consonants and vowels of this noise you're making" and even if they were to say such a thing, it would have no consequences for the learner's actual experience.* (p. 9)

Liberman (2008): *And the mystery is why HVPT [high variability phonetic training]— a simple, quick, and inexpensive technique for helping adults to learn the sounds of new languages — is not widely used. In fact, as far as I can tell, it's not used at all. Over the years, I've asked many people in the language-teaching business about this, and the answer has always been the same. It's not "Oh yes, well, we tried it and it doesn't really work;" or "It works, but the problems that it solves are not very important;" or "I'd like to, but it doesn't fit into my syllabus." Rather, their answer is some form of "What's that? I've never heard of it." But for me, the biggest question is a sociological one: why the big disconnect between research and practice?*

Both theoretical and experimental work addressing questions related to the acquisition of a first language (L1), or a second language (L2), acknowledge the importance of *input*, whether it is for learning the phonemes of a language, providing evidence of subject-verb order, or the details of the prosodic system. But when issues related to *adult* L2 acquisition emerge, the precise details, circumstances and quality of input, is the subject of much debate. Bley-Vroman (1990) presented a now well-known argument, based on the course of adult L2 acquisition and

the typical end-state grammar of adult L2 learners, that the two processes are not comparable; L1 acquisition is "fundamentally different" from adult L2 acquisition. Nonetheless, scholarly articles in experimental, applied, and theoretical research in Second Language Acquisition (SLA), regularly include a section dedicated to quantifying or qualifying the motivation, aptitude, or education of their subjects and the nature, quality or conditions of "input" (see Truscott & Smith, 2011, for an overview).

Further complicating the picture of adult SLA are questions related to *instructed* language learning. Currently, there is an emphasis on interacting with the target language with comprehensible input in meaningful and authentic contexts, along with some combination of explicit and implicit learning/teaching (DeKeyser, 1998; Ellis, 1998, 2004; Yuan & Ellis, 2003). However, there is little agreement on what should be taught, under what conditions and when (see Ellis 2005, 2006 for overview and discussion). While it is accepted that learning a language is usually with a goal of communication, pedagogy that uses the goal as the method seems to make extracting the correct form of the input/output uncertain.

Additional factors thought to impact the progression and acquisition of a target language include Length of Residence (LOR), Age of Onset (AO), socio-cultural conditions and issues related to identity, as well as self-esteem, and other affective filters (Gardner, 1985; Long, 1990; Schumann, 1975). Given the sheer number of conditions, all of which may interact on multiple levels, a cogent explanation of the general state of the L2 inter-language or end-state phonological grammar seems intractable. And yet, it seems equally implausible that not one study involving adult second language learners had a single participant that met all criteria for successful acquisition; the presence of plentiful, comprehensible and meaningful input, the learner's attention, motivation, positive attitude, adequate LOR and good self-esteem. On the

contrary, such learners do exist, but there is little evidence of end-state, native-like phonological grammars.

Given that persistent accented speech is the norm for adult learners, then the assertion that sufficient input (under optimal conditions) is adequate for successful acquisition is not supported by the state of the adult L2 phonological system. With respect to all other domains of language, the phonological system of a foreign language is more subject to age effects and immune to longer LOR (Granana & Long, 2012; Larson-Hall, 2006; McAllister, 2001). What mechanisms of adult language processing can account for this apparent difficulty?

Most models of adult L2 language processing begin with the consequences of L1 acquisition. An infant's perceptual system undergoes a re-organization within the first year of life that reflects a neural commitment (Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992; Kuhl, 1993) to whatever language they encounter in the immediate environment — presumably the L1. This re-organization is thought to make first language processing more efficient. As the instances, or exemplars, of the L1 accumulate, representations are enhanced, which decreases sensitivity to language sounds that are not part of the L1 inventory. This decrease in sensitivity coupled with a dense population of L1 speech sounds effectively filters the phonetic cues that would signal a speech sound in the targeted L2.

There are several models of cross-linguistics speech perception, discussed extensively in the relevant literature, that have been advanced to account for the difficulties associated with adult L2 learning: Best's Perceptual Assimilation Model, "PAM," (Best, 1994; Best, McRoberts, Goodell, 2001), PAM-L2, (Best & Tyler, 2011), Flege's Speech Learning Model, "SLM," (Flege, 1991), Kuhl's Natural Language Magnet Model, "NLM" (Kuhl, 1991, 1993, 2000), and more recently, Strange's Automatic Selective Attention, "ASP," (Strange, 2011). Each model

attempts to either account for, or predict, those non-native phonemes that an adult learner is likely to misperceive or miscategorize. Depending on the theoretical orientation of the scholar as well as what population of L2 perceivers they are referencing, these models differ on what feature, gesture, or phonetic detail of the L1 plays a role in obscuring the details of the L2 speech sounds (discussed further in section 2. 5). Nonetheless, the basic assumption is that the L1 phonological system interferes with the accurate perception of L2 speech sounds such that relevant input goes unnoticed and hence cannot be integrated into the burgeoning L2 grammar.

Although not explicitly considered together in the same literature, the filtering effects proposed in L2 phonological acquisition bear a striking similarity to assumptions about processes of speaker normalization in general speech perception. Normalization was proposed in order to account for one of the earliest challenges to models of speech perception; what are the identifiable, invariant features of speech that allow listeners to reliably retrieve meaning? That is, given the amount of variation in the surface phonetic forms of utterances, both within and between speakers, how can humans understand or encode speech? Speaker normalization, then, involves factoring out phonetic variation in surface forms, extracting invariant, abstract features and mapping these to a mental lexicon populated by underspecified, abstract lexical items. Normalization and models of L2 perception, lead to a similar conclusions: phonetic details available in the surface instantiation of language may be ignored in the process of mapping to abstract representations in the lexicon.

The assumptions of normalization do not adequately account for evidence suggesting that listeners are sensitive to the phonetics details of an interlocutor's speech. Evidence of such sensitivity comes from several different areas of research: evidence that allophonic variants are encoded "as is" rather than mapping to an abstract phoneme; encoding of speaker's indexical

features such that later processing is privileged when encountering the same speaker's voice; the convergence, (or in some cases divergence), on a shared phonetic space between conversational partners; and the bi-directional influence of the L2 on the L1 for new learners of an L2. Given these areas of research, there is an apparent disconnect between the general failure of adults to acquire L2 phonological categories and evidence that other listeners store phonetically-detailed exemplars in memory.

The models of L2 perception, NLM, PAM, PAM-L2, SLM, and ASP offer possible reasons for the apparent inability of adult L2 learners to accurately categorize — or to create new categories — for non-native phonemes, at least at the initial state. However, given the sheer number of exemplars a learner would encounter, for example, during a 4-year academic immersion environment, it seems inconceivable that even a biased perceptual system would not 'shift,' broaden, or become more diffuse as exemplars from the environment accumulate. If it is the case that categories are learned through experience with language such that instances are stored as a set of exemplars, then wouldn't one reasonably expect a shift in the category space, particularly of learners in immersion environments? If L1 categories are attracting L2 phonemes, and encoding them as instances of the L1 category, one would expect an eventual shift of the central tendency of the category; this shift should then subsequently begin to neutralize the 'magnetic effects' of the L1 category. Rather than pathologize all adult L2 learners, I entertain the possibility that something more than mismatched phonemic categories impact L2 phonological encoding. In the introduction to this work, I listed among the variables required by educators for successful adult L2 learning was meaningful communication in authentic contexts. I maintain that this is exactly the environment that prevents the creation of new categories. Empirical evidence from research in the neurosciences supports the claim that processing input

for *meaning* is different from processing input for acoustic details. If a search for meaning deafens learners to the phonology of the target language, how can learners listen "beyond" meaning?

There is a large body of work that treats laboratory training of non-native phonemes (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Guion & Pederson, 2007; Jamieson & Morosan, 1986, 1989; Jongman & Wade, 2007; Lively, Logan, & Pisoni, 1993, 1994; Lively, Pisoni, Yamada, Tokhura, & Yamada, 1994; Logan, Lively & Pisoni, 1991; Logan & Pruitt, 1995; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; Pisoni, Lively, & Logan, 1994; Protopapas & Calhoun, 2000). In the most general terms, training consists of repeatedly presenting minimal pairs that highlight the phonetic features that serve to distinguish targeted speech sounds. Much of the training includes a relatively small number of minimal pairs, either naturally produced by native speakers, or synthetic tokens manipulated to exaggerate the targeted dimension. In a laboratory, participants are given a forced-choice task where they must identify or discriminate one pair of a minimal pair, which differ by one sound or feature. Often they are provided immediate feedback on the accuracy of their choices. This type of training is both consistent with and reflective of exemplar models of encoding wherein the lexicon is purported to consist of "a collection of specific individual perceptual episodes or tokens" (Pisoni & Levi, 2005, p.10) over which generalizations emerge.

Phoneme training is generally successful (particularly when compared to control groups) but to my knowledge, few studies have extended such findings and methods into a classroom environment, save Wang and Munro (2004) for vowel training and Shih, Lu, Sun, Huang, and Packard (2010) for Mandarin tones.

This dissertation reports on classroom-based training of North American English (NAE) vowel contrasts [æ][1]~[ɛ], [i]~[ɪ], and [u]~[ʊ] for Chinese and Korean graduate students enrolled in an advanced pronunciation class at a university in the Midwest. Vowel training materials consisted of 486 nonce minimal pairs of the form CVC generated with onsets and codas balanced for place and manner of articulation. The nonce words were recorded by eight L1 English speakers at three speech rates and were then distributed between four training paradigms. The training paradigms were constructed to examine the effects of variation, namely, multiple speakers, different rates of speech, combinations of speaker(s) and rate(s), and coda context on vowel learning. Training was assigned as homework over the course of six days with pre- and post-tests on both real and nonce words preceding and following training.

There are several questions of interest that are pursued through this research. The first is to determine if training developed and tested under laboratory conditions can be successfully incorporated into classroom teaching. The web based application instantiated here, was straightforward, in the sense that training tokens were not synthetically generated or altered; the training did not require a laboratory and took place in the context of a regular class. In other words, conditions and tools that might be replicable and accessible in any language teaching and learning context were used for this training. That being said, given the amount of control a laboratory setting provides, it is not at all clear that similar results should be expected when relying on students to complete homework – particularly given the demands these participants faced in their respective graduate programs.

---

[1] While /æ/ patterns with lax vowels in that it cannot occur in an open syllable, Wang and Van Heuven (2006) who also treat it as a tense vowel, note that /ae/ is longer than all other lax vowels, as long as any tense vowel in the inventory and occupies the outer edges of vowel space comparable to [iː], [e],[o],[ɔ] and [u]. This categorization finds support with Shriberg & Kent (2003) and is regularly used as a tense vowel in 'tense-lax vowel studies', e.g., Chen, 2006; Wang, 2008, Tsukada et al. 2005, Norbre-Oliveira, 2007.

The second area of interest is how the L1 vowel inventory impacts the course of learning. For L2 learners who have interacted with English in both formal educational settings in their home countries and *in situ* at a U.S. university, how does the L1 vowel inventory impact initial testing, the course of learning, and performance at post-test? The literature on the vowel inventory of Mandarin reports various numbers of vowels depending on the theoretical orientation of the researcher. For this work, I reference two empirical studies, that of Wu (2011) (see also Shih, 1995) who provide both acoustic and articulatory evidence and Thomson, Nearey, and Derwing (2009), who provide production and perception evidence. Their findings, regarding the characteristics of the vowels in question and their cross linguistic similarities and differences, were generally in accord with each other. For Korean, the work of Yang (1996) who also provides acoustic information on formant frequencies for both male and female speakers is referenced. Such information is necessary since labeling vowels as equivalent, cross-linguistically, can be misleading and may result in inaccurate predictions for vowel perception and learning.

A third area of interest is the interplay between coda conditions and vowel learning. Many vowel training and/or perception studies have used consonant-vowel (CV) tokens, but as Strange (2011) notes, "in most experiments on L1 and L2 perception of consonant and vowel contrasts, the use of a single syllabic context (e.g., CV syllables) affords no way to determine whether a phonemic or systematic allophonic level of analysis is being tapped" (p.450). The training material developed for this study was of the form CVC with codas /p, t, k, b, d, g m, n, s/ balanced across vowels. Mandarin, if there is a coda at all, permits only [n] and [ŋ]. Korean has a larger number of licit codas /p/, /t/, /k/, /m/, /n/, /ŋ/, and /l/ but excludes voiced stop consonants

and /s/ from codas (Ha, Johnson, & Kuehn, 2009). Details of both vowel systems and coda conditions are expanded in Chapter 3.

A final area of interest lies with questions related to the interaction of variable training material and the contrasts to be learned. From the perception literature, there is evidence that multiple speakers and different speech rates – but not volume changes – affect participants' ability to recall spoken word lists (Goldinger, 1991; Mullennix, Pisoni, Martin, 1991; Nygaard et al., 1993; Pisoni, 1993; Sommers, Nygard, & Pisoni, 1992). From the perceptual *training* literature, there is evidence that multiple talkers results in more robust category formation; however, only Sonu, Kato, Tajima, Akahane-Yamada, and Sagisaka (2013) have used speech rate as a controlled variable in a training experiment. As such, speech rate is one element of variation under consideration here. Does variation in the training material, whether due to rate of delivery, multiple speaker voices, or coda condition, promote learning of all contrasts equally or are some vowels more difficult than others? Are some kinds of variation better than others? The research questions, then, can be summarized as follows:

1. Can experimental methods from the laboratory be employed successfully in the classroom?

2. How does the L1 vowel inventory impact the initial perception and subsequent learning of L2 vowels?

3. Do coda consonants and related co-articulation effects, impact the perception and subsequent learning of vowel contrasts?

4. Are all vowel contrasts equally learnable given more or less variation in the training materials?

These research questions, additional discussion, and predictions are expanded in Chapter 3. The following chapter, Chapter 2, provides a review of the relevant literature, beginning with the factors purported to affect adult second language acquisition. In Section 2.2, the experimental literature on adult acquisition of phonology is reviewed, which seems to indicate that the phonological grammars of adult L2 learners are resistant to change regardless of input. The general failure to acquire a native-like L2 phonological system is considered in contrast to empirical evidence that suggests that listeners are sensitive to the phonetic details of the acoustic signal and that these details are stored in long term memory. The literature detailing research supporting sensitivity to and storage of fine phonetic detail is presented in Section 2.3.

Section 2.4 features exemplar models of encoding which are presented as support for findings that listeners store phonetically detailed experiences with language. While there is ample experimental evidence to support exemplar models, the explanatory power of these same models seems compromised given the evidence of L2 phonological acquisition. Although NLM, PAM, SLM, and ASP are thought to account for apparent difficulties encoding L2 speech sounds, in Section 2.5, I review the literature that proposes that processing input for meaning is different from processing language for acoustic details, and as such, cues relevant to phonetic or phonological differences go unnoticed. In Section 2.6, laboratory training methods are reviewed that provide ways to circumvent the complications associated with constructing meaning in discourse and as a successful method for learning new, non-native phonemes. In detailing these training methods the consistencies with exemplar models of encoding are highlighted, while reviewing those training protocols determined to provide the most efficient perceptual learning. In Chapter 3, information on the vowel inventories and phonotactic constraints operating in Mandarin and Korean is presented. A discussion of how models of L2 perception, particularly

ASP and SLM, might speak to predictions regarding perception and learning are also discussed. Design considerations are elucidated in the Methods section in Chapter 4 and in Chapter 5 the results of training are presented. The results are further examined in Chapter 6. In Chapter 7, areas for future research and some reflections on using phonemic training in classroom teaching are discussed.

**Chapter 2. Literature Review**

**2.1 Adult SLA and "Input"**

In order to acquire, or learn, a language, whether it is a first, second or third, clearly some kind of interaction with the target language is necessary. This interaction is relied upon to provide *input* to the system, which in turn provides the evidence necessary for determining the phonology, syntax, morphology and semantic values of the target language. Given, however, that adult second-language learning is replete with evidence of non-target like mastery of any or all of these elements, there is extensive hypothesizing about the processing of input, and how, (or if), it is incorporated into the developing grammar of a learner (Carroll, 1999, 2001; Gass, 1997; Krashen, 1982; Van Patten, 1996, 2002).

Several theories have evolved around the processing of input in SLA. Gass (1997) for example, posits that a learner handles input in stages so that initially, input is "apperceived," but might come, at least, with some awareness that an aspect of the L2 is not part of the learner's L1 repertoire. Additionally, "comprehended input" may be noticed by the learner to the extent that the input is available for further analysis and, as such, may be stored in memory in service of the developing grammar. Krashen's (1982) hypotheses are that input is subconsciously stored in the brain in the context of meaningful communication and that affective filters, such as anxiety, can stymie learning. Communication, therefore, must be meaningful and take place in an environment that is low-stress. Carroll (2001) in the Autonomous Induction Theory asserts that speech processing is automatic and outside of conscious control so that theories "claiming that acquisition must be conscious [are] misguided. . ." (p. 227). She also asserts that that there are

multiple stages at which linguistic stimuli could impact intake, integration and output, depending on what module (phonological, syntactic, or conceptual), is involved in processing.

While input is central to much of the theorizing on SLA, it is also commonly acknowledged that the presence of input does not guarantee intake. As Corder (1967) notes,

> The simple fact of presenting a certain linguistic form to a learner in the classroom does not necessarily qualify it for the status of input, for the reason that input is "what goes in" not what is available for going in, and we may reasonably suppose that it is the learner who controls this input, or more properly this intake. ( p.165)

Van Patten (1996, 2002) emphasizes the importance of working memory and attentional resources in order to convert input to intake. He notes that learners are primarily scanning input for meaning but are expected to concomitantly attend to accurate forms in the target language. To make the bridge between input and intake, there seems to be consensus among researchers, (save for Carroll [1999, 2001]), that attention, and its various synonyms, (Posner & Petersen, 1990, "alertness, orienting, detection;" Schmidt, 1990, 2001, "consciousness" "noticing" "noticing the gap;" Gass, 1988, 1997, "apperceived input" and "comprehended input") is central to successful intake and subsequent integration of input to a developing L2 grammar. Consistent with Carroll's Autonomous Induction Theory (1999, 2001) one might ask, "attention to what?"

In many theories of teaching and learning, input is discussed in terms of the context, task and the nature of instructional material, rather than in terms of the learner's attention or understanding. For example, Bachman (1990) asserts the importance of interactional authenticity, indicating that teaching materials should be designed to deliver input in the context of plausible, real life situations. Variations on this idea are reflected in assertions of the necessity of communicative, task-based teaching (Ellis, 2003; Long & Robinson, 1998), comprehensible

input in meaningful (possibly negotiated), interactive, contexts (Krashen, 1981; Long 1983; Long, 1996), and a combination of explicit and implicit learning (DeKeyser, 1998; Ellis, 1998; Ellis, 2004; Yuan & Ellis, 2003). There is little quantitative evidence to support the assertion that such instruction provides better outcomes (Becker, 2001). Not to mention, one of the fundamental elements, that of "input," is rather elusive; it is difficult to quantify, nearly impossible to qualify, and really only constitutes input *if* the learner notices it. The hypothesis that L2 learning happens in the context of meaningful, communicative interaction is, it seems, antithetical to noticing the form of language. A discussion of differential processing of input for meaning versus form is discussed in Section 2.7.

In empirical work in SLA, notions of "input" may also suffer from an impoverished definition. Researchers usually report on some element of their participants' learning conditions, e.g. naturalistic, formal/classroom, immersion, etc. but it is rarely an experimental variable, per se. Rather, the variables Length of Residence (LOR) in the target-language environment, Age of Acquisition (AOA) , (often conflated with Age of Arrival (AOA) in the target-language country), and Age of Onset (AO) which can be used to indicate when a learner began formal learning of the target language or when the learner was first exposure to the target language, are employed as proxy indicators for the nature and amount of input. It should be noted that I do not intend this is as a criticism; merely an observation that there are no good ways to quantify input, (outside of a classroom that is taking place in a vacuum). In the following discussion of relevant research, I default to these same variables.

**2.2 Adult SLA and the Ineffectiveness of "Input"**

Rather than assuming the presence of the aforementioned optimal conditions of meaningful, comprehensible, and interactive authentic input, attention, motivation and good self-esteem, let it suffice to hypothesize the existence of some optimal condition for any given adult L2 learner. Should one expect to find an adult L2 learner who "can pass as native?" In some realms of the L2 grammar, (particularly the syntax and the lexicon), the answer could be yes. Van Patten (1996, 2002) notes, in advocating the efficacy of certain pedagogical strategies that focus on syntax, morphology and other functional items in meaningful contexts, provides the caveat that he "makes no assumptions about phonology or other aspects of the linguistic system" (p. 796). However, as Carroll (1999) asserts, "there are no motivated grounds for assuming that the acquisition of phonetic and phonological knowledge is to be accounted for *in principle* in ways distinct from accounts of the acquisition of morphosyntax" (p. 343). Nonetheless, in terms of phonology, there is a paucity of examples of end-state grammars that could be considered "native-like" even when learners have lived several decades in the target L2 environment, when years of study are involved or even when motivation is exceptional.

Goto (1971), for example, found that even among advanced Japanese learners of English with relatively good pronunciation, accurate perception of /l/ and /r/ was well below that of native speakers. He postulated that those with better pronunciation were using "kinetic sensations of their own speech organs as deaf persons would do" (p. 321) and that despite this adaptation with their articulators, they were still unable to make use of the auditory feedback from their own productions. MacKain, Best, & Strange (1981) found improved discrimination but only with learners who had had language coaches for 8 hours a day for well over a year. Riney and Flege (1998) tested Japanese speakers at an American University for both global

accent and /l/~/r/ identification at their first year as freshmen and then again in their senior year. Despite a 4-year LOR, no significant difference was observed for the majority of the participants, although "some speakers made significant improvement… in both global foreign accent and liquid identifiability and accuracy" (p. 213).

In an experiment replicating that of Flege, Takagi, and Mann (1995), Larson-Hall (2006) tested 30 L1 Japanese participants, 15 of whom had lived in the U.S. for less than three years (the "inexperienced" group) and 15 of whom had resided in the U.S. for twelve years or more (the "experienced" group). Both groups were recorded first reading a story and then reading a word list both of which were then rated by a panel of L1 American English judges for accentedness and accuracy. Larson-Hall found that for the experienced speakers, there was a negative correlation with accentedness ratings at both the word and sentence level. In other words, longer residency did not improve learners' ability to accurately produce /l/ and /r/, rather, she found the opposite; that accuracy declined with longer LOR.

In a similar finding to that of Larson-Hall, McAllister (2001) found no significant difference in perception of Swedish vowel contrasts between learners with short LOR (3.6 years) and longer LORs (18 years). Although Granana and Long (2012) were interested in a combination of factors, Age of Onset (AO), LOR and aptitude, they controlled for LOR such that participants (65 L1-speakers of Chinese) had all lived in Spain between 8 to 31 years. As they note in their findings "the decline in pronunciation started very early and was already significant in the group that had started learning at 3 to 6 years of age." In summary, of the learners they examined, none with an AO later than 5 years-of-age were judged to sound native-like.

There are a lot of assumptions at work in implying that LOR provides some measure of the availability of input; it is, after all, possible to live in an L2 environment and interact

predominantly in the L1. Given the amount of research on L2 acquisition and the number of participants, coupled with the fact that so much of L2 research takes place on or near university campuses, it would not seem a Pollyanna-like expectation that there should be many L2 speakers benefitting from optimal learning conditions. And yet, most studies find little evidence of L2 learners with a phonological grammar approaching L1 norms. Perhaps it is telling that the few researchers examining instructed L2 phonology are calling for research on pedagogical methods that target *intelligible* (rather than native-like) pronunciation (Saito, 2012).

## 2.3 Linguistic Experience and Phonetic Perception

How does one account for adult second language learners' apparent difficulty acquiring the phonology and phonetic details of a target language? One part of a possible answer is the effects of early experience with language. Kuhl et al., (1992) provided evidence that infants, at birth, are sensitive to the phonetic details of all speech sounds, regardless of L1. At six months of age, however, infants begin homing in on the ambient language and thus demonstrate a decreasing sensitivity to cross-linguistic phonetic differences. This preference for the L1 reflects a neural realignment, or as Kuhl (2004) puts it, "a neural commitment," to the ambient language and is thought to facilitate learning words and other more complex structures (Kuhl, 2004; Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson, 2006).

In her 2006 study, Kuhl et al. demonstrated a facilitative effect for L1 processing at the expense of a more universal discriminatory sensitivity. In this experiment, she tested the discrimination of the /r/~/l/ contrast on infants born into an American English-speaking environment and infants born into an L1 Japanese speaking environment. Before six months of age, the two groups of infants were equally poor, compared to L1 adult perception, at

discriminating /l/ and /r/. By seven months, the American English infants improved significantly at /l/~/r/ discrimination, while the Japanese infants' ability had *declined*. These effects were also shown to correlate with later language abilities measured at 14, 18, 24, and 30 months of age. Infants who attuned to the L1 earliest, thereby losing the ability to perceive non-native contrasts, experienced accelerated language development. Those infants who were better at discriminating non-native contrasts remained in the "initial universal stage of phonetic perception" (p. F19), uncommitted, which is reflected in slower language growth, again as measured at 14, 18, 24, and 30 months of age. The findings from this body of research suggest that early language experience with the phonetic units of the native-language results in a neural commitment that facilitates processing of the L1 and "plays a role in the decline of nonnative phonetic perception" (Kuhl, 2006, p. F19).

These findings might be considered, however, in light of Pierrehumbert's 2003 treatise on statistical learning and category formation where she reflects on the amount of experience necessary to master a phonological category. Referencing research by Hazan and Barrett (2000) on the gradient perception of categories by children aged 6 – 12 years of age, Pierrehumbert notes that "some parts of the speech processing system are initiated early; however, the system takes a long time to develop, not achieving adult levels even at twelve years of age. . ." (p. 115). In other words, the developing phonological grammar is not complete and fixed at six months of age.

**2.4 Linguistic Experience and Models of L2 Perception**

In response to Kuhl's research, several models of adult L2 speech perception have found purchase in the SLA research literature. Directly from her infant studies, Kuhl's Natural Language Magnet model, "NLM" (Kuhl, 1991, 1993, 2000) was created to model the neural response of infants' perceptual systems from their early experiences with the L1. In research reported in Kuhl's 1991 article, she considers the internal structure of categories in adults, infants and monkeys. She found that for both adults and children (although not monkeys) certain exemplars of the category /i/ were perceived as better exemplars than others. For example, if a perfect exemplar, a prototype, of /i/ was first presented to the participant, a second /i/, which was a certain psychophysical distance from the prototype, might be judged as another "good" exemplar; however, if the second /i/ from the preceding example were presented first, followed by a third /i/ that was equal in psychophysical distance from the second and still a category member (but further from the prototype), the third /i/ would be judged a poor exemplar of the category. Good fitting members clustered around a particular region of a category space and lead Kuhl to hypothesize a perceptual magnet effect around category prototypes.

Subsequent work (Iverson, Kuhl, Akahane-Yamada, et al., 2002) has provided additional evidence of the effects of L1 experience wherein participants who were not learners of English but rather monolinguals living in the home-countries of their respective L1s perceived NAE /l/ and /r/ in a manner consistent with their respective L1s, e.g., German, Japanese, and NAE.

Best's Perceptual Assimilation Model, "PAM," (Best, 1994; Best et al., 2001) is cited in numerous studies examining perception of non-native speech contrasts but as she clarifies in Best and Tyler (2011):

We define nonnative listeners more strictly than has often been done in the past. For us, they are functional monolinguals, i.e. not actively learning or using an L2, and are linguistically naive to the target language of the test stimuli. By comparison, L2 learners are people who are in the process of actively learning an L2 to achieve functional, communicative goals, that is, not merely in a classroom for satisfaction of educational requirements. (p.16)

In PAM-L2, Best and Tyler extend the model to account for L2 learner but as they say explicitly, their interests lie not in laboratory training or in classroom learning, their "interest lies instead with natural communicative situations, which more broadly engage the multi-tiered grammatical and phonological structure of the L2. This is more consistent with theories of second language acquisition (SLA) that hypothesizes meaningful conversation is the main context in which the properties of a new language are learned (e.g., Carroll, 1999)" (p. 19).

Flege's (1991) Speech Learning Model, "SLM," was intended to model the perception and production of adults actively engaged in learning the L2 and operates at the level of individual phones (rather than contrastive pairs). SLM posits that perceived phonetic equivalence or difference will facilitate or inhibit learning of target phones, predicting that two sounds that are similar but not identical are the most difficult to learn and that the more dissimilar two sounds are, the easier it is to notice the differences and thus learn a new category. Measuring, however, the perceived differences of two sounds is difficult and particularly so given that SLM also argues that the perception of an L2 phone will depend on the context in which it occurs, for example, an [ɛ] may be perceived as a member of a different category when it is followed by /ŋ/ and when it is follow by a /b/.

Strange (2011) proposes a model, Automatic Selective Attention (ASP), such that two modes of perception, phonological or phonetic, interact with task complexity, listener's knowledge, and the allocation of attentional and cognitive resources. Input is further subdivided according to two types of contrast saliency; auditory salience and perceptual salience. Phonological perception mode means that a listener ignores much of the variation in the signal and engages bottom-up processes in order to recognize words via "highly over-learned, automatic selective perception routines (SPR)" (p. 460). As such, it is fast and efficient and attention is just enough to retrieve details to extract meaning. Phonetic perception mode means that listeners are not engaged in recognizing words in connected speech, for example, rather they are attuned to phonetic information including that related to allophonic variants as well as phonotactic patterns of the L1. These modes of perception also function in the context of listening to an L2. In terms of the two types of contrast saliency, auditory salience describes perceptible differences in stimuli presented to a naïve listener in, for example, a perception study. Perceptual salience, on the other hand, is used to describe contrastive stimuli presented to L2 learners, actively engaged with learning the L2, for whom the L1 may impact their ability to discriminate differences.

In this thesis, I am not undertaking the task of testing one model against another; rather, I wish to establish that some models, e.g., PAM and NLM are not appropriate for training research involving experienced L2 learners such as is presented here. While SLM clearly pertains to the population of this study, it is not necessarily obvious how one is to establish similarity between sounds or even what constitutes a "new sound." For Strange's ASP, if it is assumed, given the nature of the training task, that learners are in "phonetic perception mode," how would one then ascertain whether the stimuli are provoking reactions based on perceptual or acoustic saliency?

These questions will be revisited in Chapter 3 where the vowel inventory of Mandarin and Korean are presented in order to make several predictions vis-à-vis the initial state of their perception of NAE vowels as well as the effects of training.

In summary, early experience with an L1 shapes the perceptual organization of the listener; difficulties with perceiving the phonetic details that categorize phones in the L2 are impacted by this early organization. The L1, in essence, acts as a filter through which the L2 is perceived.

**2.5 Speaker Normalization**

The filtering effects inherent in the models of L2 speech perception have a counterpart in general models of speech perception. A long-standing quandary in speech perception is how we understand speech given the amount of variation evinced across speakers. For example, in the well-known Peterson and Barney (1952) study of vowel formant frequencies, 33 men, 28 women and 15 children were record producing two repetitions of ten vowels in the same consonant context (hVd). Hillenbrand and Gayvert (1993) revisit the Peterson and Barney data in their study addressing questions related to vowel identifiability. They note that there was "considerable formant-frequency variability from one speaker to the next, and there was a substantial degree of overlap in the formant-frequency patterns among adjacent vowels," nonetheless, they write, "The vowels were highly identifiable" (p. 668). The Peterson and Barney graph (Fig.2-1), with clouds of overlapping vowels, is one of the earliest visual representations of the amount of variation in the surface forms of utterances. It is a representation that begs the question: How is it possible for interlocutors to retrieve meaning from the speech stream when there are apparently no identifiable, invariant features?
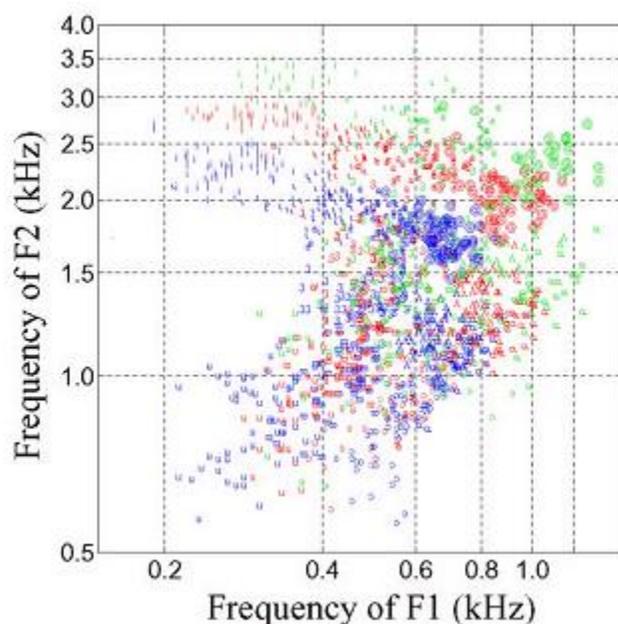
Fig. 2-1 Peterson and Barney (1952)

Speaker normalization is the proposition that superfluous variation in the acoustic signal due to dialectical, physical or contextual differences is filtered out by the listener and abstract features are retrieved, processed, and matched to abstract lexical representations (Cornell, Lahiri, & Eulitz, 2011). As illustration, assume a listener from the Peterson and Barney (1952) study, who heard the word "heed" 152 times produced by 76 different people, each of whom said the word "heed" differently from that of any other person. A speaker normalization process would allow the listener to perceive the correct word 94.4% of the time, regardless of whether a small child or large man had produced it. Ames and Grossberg (2008) assert that, "speaker normalization helps the brain to overcome a combinatorial explosion that would otherwise occur if the brain needed to store every instance of every speaker utterance in order to understand language meaning" (p. 3918). It is, as Goldinger (1998) observes, "information reduction in the

interest of economy, a process of data abstraction that yields a minimal, symbolic representation from analog inputs" (p. 251).

Normalization and models of L2 perception, then, have similar assumptions: phonetic details available in the surface instantiation of language are largely ignored. Normalized or filtered, it is perhaps understandable then that a North American English speaking student might fail to notice the phonetic difference between a Spanish /t/ and an American English /t/. That being said, there is a growing body of evidence that listeners are not only sensitive to the phonetic particulars of speech but that these details are remembered. The tenets of normalization are inadequate to account for such empirical evidence.

## 2.6 Storing Phonetic Details

Countering models that require listener normalization is a proliferation of experimental data that seem to indicate that listeners are both sensitive to and store in long term memory phonetic details. Such evidence comes from areas of research addressing the encoding of allophones, the encoding and storage of indexical features of voice, interlocutor convergence or divergence, as well as the bi-directional influence of the L2 on the L1.

In experiments with North American English (NAE) listeners, Connine (2004) found evidence that her participants were more likely to hear *pretty* over a non-word *bretty* when presented with word- non-word sequences of sounds that contained an intervocalic voiced alveolar flap. The flap is an environmentally conditioned allophone of /t/ in NAE that occurs in an intervocalic position, preceding an unstressed syllable. In this position, it is more common for NAE speakers than the voiceless stop [t]. Given her outcomes, Connine concludes that listeners weren't accessing a single abstract phone, stripped of all predictable variation, "listeners do not

*recode* the flap variant into an underlying /t/ version but recognize the flap, in its surface form, via a preexisting representation in lexical memory" (p. 1084). Connine asserts that this is inconsistent with a model of word recognition in which *pretty* has a single underlying phonological representation with a voiceless alveolar stop, which is recovered by the listener from the flapped variant.

Another example of listeners' sensitivity to phonetic detail that calls into question processes associated with normalization is evidence suggesting that indexical features related to the identity of a speaker are encoded in long term memory. In this body of research, listeners are first familiarized with the voices of novel speakers and then charged with a linguistic processing task such as word recognition, list recall or sentence transcription. Listeners evince a "Familiar Talker Advantage" such that later processing is advantaged when a familiar talker presents the task. These studies (Goldinger, Pisoni, Logan, 1991; Magnuson & Nussbaum, 2007; Nygaard & Pisoni, 1998; Palmeri, Golinger & Pisoni, 1993; Pisoni, 1993; Nygaard, Sommers, & Pisoni,1994; Schacter & Church, 1992) have consistently shown that listeners' performance on linguistic tasks is consistently better for familiar talkers than for unfamiliar, novel talkers. In order to develop such advantages, a listener must perceive and remember the phonetic details attributable to a specific speaker.

Research that provides evidence of listener sensitivity to phonetic detail is present in the literature that treats unconscious and automatic synchronization (also called convergence, imitation, accommodation or entrainment) of speech during the perception-production loop. Entrainment and behavioral synchrony have been studied in a variety of disciplines, including music perception, dance, verbal communication and motor coordination more generally. In the realm of verbal communication it has been demonstrated that interlocutors converge on a shared

discourse space. In other words, speakers engaged in conversation, for example, may repeat each other's lexical choices or use similar syntactic constructions (Garrod & Anderson, 1987). In Putman and Street (1984), participants, under a "likeable" condition, converged on speech rate, response latencies and turn durations. At another level of phonetic detail, convergence has been found for the formant structure of vowels (de Boer, 2000); voice onset time (VOT) (Nielson, 2008) and intonation (Braun, Kochanski, Grabe, & Rosner, 2006; Cole & Shattuck-Hufnagel, 2011; Peschke, Ziegler, Kappes, & Baumgaertner, 2009). The synching or convergence effect in the realm of speech has been observed under a number of experimental paradigms, including: cooperative, task-based language production (Cole & Shattuck-Hufnagel, 2011), imitation tasks of both real and nonce words (Namy, Nygaard, & Sauerteig, 2002; Goldinger, 1998), self-imitation tasks (Braun et al., 2006), and shadowing tasks (Marsen & Wilson, 1973, 1984).

A final area of interest regarding sensitivity to phonetic detail comes from Chang (2010). Most studies of L2 acquisition assume some interference from the L1; Chang's dissertation reports on the influence of the L2 phonology on the L1 phonology. His argument begins with evidence that adult L1 speech is malleable and that speakers rapidly adapt their productions in the face of auditory feedback, for instance, in the productions of post-linguistically deafened participants with and without cochlear implants. Given this malleability, he predicts that adult, non-proficient, late-learners of an L2 will demonstrate "phonetic drift" in the L1. Research with *highly* proficient bilingual speakers has demonstrated that there is a two-way interaction between the phonological systems, but little work has looked at the effects on novice learners. Chang's longitudinal study follows L1 English speakers in an intensive elementary Korean class during the course of which they were recorded each week. Chang found that his subjects' English productions were rapidly influenced by their experience with Korean and that the effect was one

that moved the phonetic properties of their L1-English toward Korean. This seems a particularly curious finding given that models of L2 speech perception are based on the inability to perceive the phonetic details of the L2.

Research demonstrating that detailed phonetic information is available to the listener cannot be accounted for in a model that relies on filtering out extraneous, non-meaningful variation in the speech signal in favor of an abstract prototype. However, there seems to be an apparent disconnect between the general failure of adults to acquire L2 phonological categories and evidence that other listeners perceive and store phonetically-detailed speech, which rapidly impacts perception and their own production. In the next section, an alternative model of speech perception is reviewed. Exemplar Models of language perception and encoding propose a perceptual space in which fine phonetic details are retained rather than filtered out in favor of abstract representations.

**2.7 Exemplar Models**

The hypothesis of normalization rests on the assumption that speakers filter out variation attributed to speaker, speech rate, social context and dialect. Stripped or normalized tokens can then be identified according to abstract, invariant, and category prototypes. However, as elucidated in Section 2.5, there is ample evidence that speakers are more sensitive to fine phonetic detail than a normalization process would suggest. In exemplar phonology, no single, invariant abstract representation is proposed, rather phonetically detailed experiences with language, exemplars, are remembered (Bybee, 2002; Johnson, 1997; Pierrehumbert, 2001, 2003). An exemplar then, is a detailed perceptual memory and new incoming exemplars are classified according to their similarity with previously stored items. Exemplars that are more frequent, or

more recent, have a higher activation levels. For example, if you live in Pittsburgh, Pennsylvania, activation levels for the word "pop" might be stronger than for the word "soda" (Vaux, 2003). As exemplars are committed to memory, accumulations in certain areas of a cognitive map become denser with items that are frequently encountered. Relevant for the discussion here is that the exemplar space is constantly updated, or incrementally modified, as new exemplars accumulate.

Although Bybee (2002) was discussing the paths of sound change in terms of Exemplar Theory, it is relevant for considerations in L2 acquisition. She notes that accumulations of new, frequently encountered exemplars paired with the memory decay of less-used exemplars may change the form (in terms of density distribution) of the cognitive map:

Changes in the phonetic range of the exemplar cluster may also take place as language is used and new tokens of words are experienced. Thus the range of phonetic variation of a word can gradually change over time, allowing a phonetically gradual sound change to affect different words at different rates. (p. 71)

The assumption that people learn phonological categories by remembering many labeled auditory tokens explains the ability to learn fine phonetic patterns of a language. It also explains why patterns are incrementally modified over long periods of time in adult speech (Bod & Cochron, 2007; Bybee, 2002; Pierrehumbert, 2001); with the exception, evidently, of adult L2 learners. Since learning a new sound system can be understood, in terms of exemplar theory, as the "acquisition of a large number of memory traces of experiences" (Bod & Cochran, 2007, p.2), then it could be posited that L2 learners just have not acquired a large enough number of

exemplars. While obviously there is no specific number of exemplars that correlates with the creation of a new category, in Riney and Flege's (1998) work, the L1-Japanese students were in a university setting for four years; McAllister's (2001) subjects ranged from 3.6 years to18 years living in the target-language environment and Granena & Long's (2012) participants ranged from 8 to 31 years in-country. If we assume 16,000 words per day (Mehl, Vazire,Ramírez-Esparza, Slatcher, & Pennebaker, 2007) for college students, at the end of 4 years Riney and Flege's participants would have encountered 23,360,000 words — not including advertising. Given the claims of exemplar theory (that listeners store phonetic details of everything from voice to VOT), why is it so difficult to change the perceptual map of L2 learners? Arguments for critical or sensitive periods for language acquisition have it that there is an age past which the ability to learn a new language and to discriminate non-native phonemes, deteriorates: "somewhere between the ages of 6-7 and 16-17,everybody loses the mental equipment required for the abstract patterns underlying a human language" (DeKeyser, 2000, p. 518). My assertion is that processing for meaning inhibits the encoding of new speech sounds; therefore, teaching an L2 through strictly meaningful communication is counter-productive for acquiring the target-sounds of an L2. Section 2.7 reviews some of the empirical evidence that supports the claim that processing input for *meaning* is different from processing input for acoustic details.

**2.8 Listening for Meaning**

Empirical evidence from several disciplines supports the claim that processing input for *meaning* is different from processing input for acoustic details. Carroll (2004) asserts (in response to claims of a general decline in auditory processing mechanisms of adult learners) that adults are able to easily discriminate acoustic differences when asked if one sound is different

from or the same as another sound. But, according to Carroll, when adults are asked to discriminate sounds in the context of words, "perceptual acuity is affected by the status of the acoustic distinction in the organisation of the lexical system," (p. 230). This is supported in the work of Guion and Pederson (2007) whose study on the role of attention in learning Hindi contrasts provides an example: One group was instructed that they were learning new words in Hindi (meaning-attending group) and that after completing the training, they would complete a vocabulary test in which they would pick the English word that corresponded with the learned Hindi vocabulary item. The other group (sound-attending group) was told explicitly that they were learning to discriminate the differences in the onset consonant that distinguish minimal pairs. The directions were as follows:

> Your task is to listen carefully to the beginning of each word and try to learn the difference between the Hindi sounds. You will notice that words with different meanings often have similar beginning sounds. However, the fact that they have different meanings, tells us that they are distinctive sounds in Hindi. During the course of this session, try to learn to distinguish between these sounds as best as you can. (p. 16)

In the post-test discrimination test, the sound-attending group performed better than those who were engaged in learning new vocabulary items.

Trofimovich (2008) explored under what circumstances L2 learners were sensitive to L2 phonological information. Training material consisted of 72 common English words recorded by six adult native speakers of North American English. Trofimovich divided 52 native Chinese speakers into two proficiency levels based on LOR and a spoken pretest score. The two groups of learners were then assigned to one of two training conditions. In the semantic condition, participants' attention was directed toward meaning-based processing, where they were asked to

rate the pleasantness of the word on a seven point scale (1=meaning is unpleasant to 7=meaning is pleasant). "In the control condition, no attentional orientation was imposed. However, to make this condition as comparable as possible to the semantic condition, the participants were asked to track the presentation order for each spoken word in a list. Upon hearing each word, they circled the number (from 1 to 46) that corresponded to each word's serial position on a list" (p. 317). A facilitative effect was found for familiar words in familiar voices for those with longer LOR in the semantic processing condition; that is, listeners with more experience could attend to both meaning and the phonetic details of voice which helped with later processing tasks. Those with shorter LORs were less able to encode both meaning and phonetic details of voices. Attending to meaning inhibited later processing for those participants with less language experience.

Werkers and Tees (1984) conducted four related experiments in an effort to tease apart whether adults lose the ability to discriminate universal phonetic boundaries after a critical or sensitive period or if apparent differences in perceptual acumen were due to changes in attention and/or processing strategies. At the conclusion of their experiments, they asserted that adults maintain the ability to discriminate non-native phonemes, but strategies for processing natural language inhibit the ability to detect differences. When processing speech, listeners search for meaning. They are "unable to volitionally relinquish this strategy" (p. 1866) and are therefore constrained to encoding L2 speech sounds according to L1 phonemic categories. They provide, however, that in conditions of low processing loads, they may be free to adopt non-phonemic processing.

Hickok and Poeppel (2004) propose a model of language processing that involves separate auditory-conceptual and auditory-motor processing. Based on evidence from patients with aphasia, they found that the ability to identify phonemes, rhymes, or syllables is dissociated

from the ability to comprehend words. They assert that in normal conversational speech, "listeners have no explicit awareness of the phonemic structure of the input, only the semantic content. Clearly, then, explicit access to sub-lexical structure entails cognitive mechanisms that are not typically engaged in listening to speech for comprehension" (p. 76).

Phonological processing as a whole has been shown to occupy different regions of the brain from, for example, semantic processing (Cousin, Peyrin, Pichat, Lamalle, Le Bas, & Baciu, 2007) with an even more pronounced effect in non-native speakers (Pillai, Araque, Allison, Sethuraman, Loring, Thiruvaiyaru, Ison, Balan, & Lavin, 2003). The tension between evidence that adult listeners are sensitive to the phonetic details of speech and that adult L2 learners are, evidently, insensitive to the phonetic details of speech leads to the hypothesis that processing speech for meaning, with a concomitant increase in processing load is different than processing for phonemic information. If a search for meaning deafens learners to the phonology of the target language, how does one cut through the noise of meaningful speech? The next section provides a review of the literature on methods developed in laboratory settings for training non-native contrasts. These training paradigms were designed to address listeners inability to perceive (or more accurately "to correctly classify") particular elements of the targeted phonological grammar.

## 2.9 Perceptual Training of Non-Native Contrasts

Perceptual training was initially developed and launched by clinicians in the speech and hearing sciences. Hodson and Paden (1983, 1991) developed a therapy procedure called auditory bombardment, which provided repeated, systematic exposure to multiple exemplars of phonological targets and contrasts (Bowen, 1998). Researchers in speech perception using

similar methods do not explicitly present their work in terms of exemplar encoding, but it is increasing acknowledged that such training is the practical extension of the theoretical model. Kingston (2003) writes: "the benefits of the training procedures implemented by Pisoni and his colleagues are consistent with Nosofky's (1986) attention weighted exemplar model of category learning. As in the training, listeners learn categories through experience with exemplars, such that new tokens are categorized according to their aggregate similarity to all previous tokens" (p. 5). Bradlow (2008) echoes the relationship wherein "item-specific acoustic-phonetic variability is encoded in the cognitive representation of experienced speech samples" (p. 302), or more simply, experiences with speech are encoded with the phonetic details intact.

Perceptual training arose from research in perception which, at its inception, was challenging models that proposed "normalization" as a way to compensate for variation in the speech signal associated with speakers' physical, social or emotional differences. Rather than filtering out variation, empirical evidence from perception experiments was leading researchers to the conclusion that the phonetic details of the speech signal were being retained in memory and that such memories could facilitate or debilitate speech processing.

For example, Martin, Mullennix, Pisoni, and Summers (1989) used a serial recall task to test the effect of speaker variability on memory. They found that participants who were presented word lists produced by multiple talkers had more difficulty recalling the ordered list of words than those participants who heard a single speaker's voice. Martin, et al. (1989) proposed that listeners who had the task of hearing and encoding multiple speakers' voices had fewer working memory resources for concomitantly encoding the ordered word lists.

Goldinger, Pisoni, and Logan (1991) used the same serial recall task but introduced different speech rates, fast, medium, and slow with corresponding time intervals between items -

that is, fast items with short intervals and slower items with longer intervals. At faster rates, participants had more difficulty with multiple speakers than with a single speaker; however, at slower rates, participants were better at list recall in the multiple-speaker condition than for the single-speaker condition. This was taken to indicate that encoding both speaker and linguistic item required more time but that it facilitated performance on the serial recall task.

Palmeri, Goldinger, and Pisoni (1993) presented word lists produced by a single speaker or by multiple speakers. Participants were asked to identify words in the list as "new", meaning that it was the first time they were hearing the word, or "old," meaning that they had heard the word before. Participants were faster at identifying repeated words when they were produced by the same speaker (the same speaker who had produced the earlier token), than when a repeated word was produced by a different speaker.

Finally, Nygaard, Sommers, and Pisoni (1995) investigated the effects of multiple speakers, speaking rate, and amplitude on serial recall. Participants heard lists of words produced by multiple speakers or by a single speaker. Each condition was further divided into participants who heard fast, medium, and slow delivery rates and participants who heard word lists presented at different volumes. As in Goldinger et al. (1991), participants had more difficulty at list recall when they listened to multiple talkers at faster rates, but performed better on multiple speakers at slower rates. Performance at the slow rate was equivalent between the multi-speaker and uni-speaker groups. There was no effect, however, for changes in volume.

Research such as the foregoing provided evidence that adult listeners were, automatically, encoding phonetic detail. Many of these same researchers were the earliest researchers in cross-linguistic speech perception and the subsequent evolution to perceptual training.

### 2.9.1 Perceptual Training

The bulk of the perceptual training literature reports on the training of /l/ and /r/ (Bradlow, 2008, Bradlow, et al., 1997; Jamieson & Morosan, 1986, 1989; Lively et al., 1993, 1994; Lively, Pisoni, Yamada, Tokhura, & Yamada, 1994; Logan et al., 199; Logan & Pruitt, 1995; McClelland, Fiez, & McCandliss, 2002; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; Pisoni, Lively, & Logan, 1994; Protopapas & Calhoun, 2000). The persistent difficulty that Japanese speakers have with this contrast has made it a particularly fertile landscape for testing the efficacy of different training variables for affecting change in the learner's perceptual space.

Although I review several training methods, in general, the training takes place over the course of several days, ranging from six days to 45, and averaging 15 days. The training material consists of relatively few minimal pairs, either naturally produced by one or multiple speakers, or one minimal pair synthetically created and manipulated to exaggerate the feature that should cue the contrast for the listener. Listeners are asked to identify, or in some paradigms, discriminate, the tokens that they hear. Discrimination tasks ask the listener to indicate if two tokens are the same or different; identification requires that the listener indicate what a sound is, according to some established labeling procedure. Most training paradigms incorporate some form of feedback, indicating to the participant if the selection that they made was accurate. The training is thought to help learners ignore irrelevant variation in the signal while concomitantly learning to attend to those phonetic clues that signal category membership. Despite differences in the training paradigms, in general, learners improve in their ability to identify the target contrast. Such research, then, demonstrates that under laboratory conditions, adult learners can learn to perceive non-native phonemic contrasts. Different training experiments have tried to isolate

different variables that impact the efficacy of training. The following section reviews the training literature with a focus on those protocols thought to impact learning. Some protocols have been categorically established as beneficial; for those protocols for which some questions remain, discussion with the relevant supporting literature is provided under the relevant heading.

### 2.9.2 Variability

In perception experiments, greater variability, in terms of the number of speakers or speaking rate, resulted in more difficulty in list recall tasks. It may seem counterintuitive then Pisoni's (1993) observation that "stimulus variability is useful in perceptual learning of complex multidimensional categories like speech because it serves to make the mental representations extremely robust over different acoustic transformations such as talker, phonetic environment and speaking rate" (p.7).

In Logan et al. (1991), six Japanese L1 speakers were trained on 68 English minimal pairs featuring the /l/~/r/ contrast in a variety of positions and produced by four men and two women. The training method, dubbed high variability training, was followed by a similar study (Lively et al., 1993) using the same tokens but produced by a single speaker. The results of the latter were such that participants, "developed talker-specific, context dependent representations" which "failed to generalize to tokens produced by a new talker" (p. 1242). These findings spawned a series of related studies all using high variability training. Lively et al. (1994) demonstrated that such training supported long term retention of /l/ /r/ perceptual discrimination abilities and Bradlow et al. (1997) provided evidence that *production* of /r/ and /l/ also improved. Taken together, it was concluded that high variable phonetic training (HVPT) facilitated robust, long-term changes to the trainee's phonetic perception (and production).

It should be noted however, that there were limitations to these studies; performance on /r/ /l/ identification varied, persistently, depending on the phoneme's position within the word such that "perceptual learning appeared to be local and required treatment at a context-dependent level" (Wang, 2008) and in several studies, (Lively et al., 1993; Lively et al., 1994; Logan et al., 1991) learners evinced a bias for familiar speakers which persisted even after several months. As Wang (2008) notes, "part of the problem might be related to the way the training tokens were presented in the training sessions. Participants listened to tokens produced by a single speaker in each training session" (p. 42). So while the training consisted of multiple speakers, they were isolated to one different speaker per day. Nonetheless, HVPT has become the standard method for research involving perceptual training.

There are, however, dissenters. Perrachione, Lee, Ha, & Wong (2011), in their research on individual variation as predictors for learning outcomes, found differences depending on initial base-line performance. His team trained participants on using lexical tone to identify simulated foreign words under two paradigms, one with highly variable training tokens and the other with low variability. Prior to training, participants were tested for pitch-perception abilities; subjects with strong abilities were divided between the two paradigms, as were those with weaker abilities. Their findings were such that high variability training enhanced learning only for those participants who had high pre-test scores, that is, those who were already quite good at perceiving pitch.

Additionally, Shih et al. (2010) found in Mandarin tone training that students improved more with low variability training rather than high variability, random order training, with the most efficient training accomplished with adaptive training (scaffolding learners from simpler to more complex tokens).

Jongman and Wade (2007) found that more difficult, highly confusable contrasts – such as those inherent in vowel categorization – benefitted from less variable phonetic training. Training routines that used prototypical examples of a vowel contrasts were more effective, for example, in training highly confusable non-native [i] which was learned best under minimally variable exposure.

While there is evidence in some of the training literature that high variability stimuli is beneficial to phoneme learning, (in particular, /l/ and /r/) there is at least some evidence that it may not be the most effective training condition for all speakers or for all contrasts.

### 2.9.3 Perceptible/salient input

McCandliss, Fiez, Protopapas, Conway, and McClelland, (2002) are proponents of Hebbian learning; each time a neuron fires and excites an adjacent neuron, it strengthens the neural pathway. In this case, each time a Japanese L2-English speaker hears [l] or [r] and maps it to the closest Japanese phoneme, the flap [ɾ], the extant perceptual category is reinforced. It follows, then, that if the Japanese participants cannot hear the [l] [r] contrast, then training only serves to reinforce the mismatched phonemes. In light of this, McCandliss and his colleagues developed sets of speech continua such that each instance of /rock/ or /lock/, for example, spanned a range of highly exaggerated tokens to more natural sounding tokens, through which participants were scaffolded. Their results were such that all of the subjects showed substantial improvement. For additional experimentation on the effectiveness of exaggerated stimuli see Protopapas & Calhoun (2000).

Although McCandliss et al. (2002) initially conducted their training without feedback[2], in a follow up study they found significant gains for their participants when they included feedback. Additionally, Anderson (1990) argues that feedback is critical for the development of any type of skill. Feedback is important because it enables subjects to determine whether what they are doing is accurate or not, and in the latter case, trainees can modify responses to stimuli (Strange, 1997).

### 2.9.4 Attention

Strange (2011) provides a brief discussion of the various ways in which researchers have used the term a*ttention* and concludes by specifying three categories; *attentional focus* "…is characterized as the volitional allocation of cognitive effort or resources to extract certain kinds of information from the input… " (p. 459). A*utomatic processes* do not require attentional focus. Examples of the latter are divided attention tasks where, for example, "attention can be directed to auditory *vs* visual input" (p. 458). Finally, *selective attention* she defines as in terms of cue weighting, e.g. attending more to durational cues than spectral cues in vowel identification. An example of *automatic processes*, may be illustrated in Gordon, Eberhardt, and Rueckl (1993) who conducted a series of four experiments exploring the role of attention in identifying phonemes. They found that when participants were distracted, solving math problems, for example, while at the same time trying to identify speech sounds, the participants' ability to rely on primary phonetic cues for differentiation decreased, that is, participants relied more on durational differences for [i] and [ɪ] (rather than spectral differences) under conditions of low attention.

---

*2*  They initially argued that, according to models of Hebbian learning, feedback should not be necessary.

An example of *attentional focus*, is provided in Guidon and Pederson (2007) who

manipulated the attention of English monolinguals engaged in training of unfamiliar phonetic

categories in Hindi. One group was instructed to attend to sound meaning correspondences and

the other group was to attend to phonetic cues that signaled one sound or another. The results

indicated that those trainees who were instructed to attend to the phonetic cues performed better

in a post training discrimination task. Guidon and Pederson felt this indicated that explicit

directing of attention resulted in better learning of novel phonemes.

The notion of attentional focus from Strange (2011) also harkens to the observation made

in Pisoni et al. (1982) in the context of experiments in speech perception, "when the

experimental conditions are modified to reduce uncertainty or when the subject's attention is

explicitly directed to the acoustic differences between the test signals rather than their phonetic

qualities, subjects can accurately discriminate very small differences in VOT" ( p. 298).

For this training study, trainees are explicitly directed to turn *attentional focus* to the

kinds of information necessary to identify vowels.


### 2.9.5 Long-term versus short-term training

Short-term training, e.g. one session, has been demonstrated to afford participants some

gains in their ability to discriminate non-native sounds (e.g. McClaskey, Pisoni, & Carrell 1983;

Pisoni et al., 1982). However, training that takes place over the course of several sessions seems

to be more effective in training robust novel phonetic categories. In most studies, researchers

provide one session per day, with the training lasting anywhere from six to 45 days. The average

is around 15 training sessions taking place over the course of three weeks. Logan and Pruitt

(1995) cite practical considerations such as those of maintaining the attention and interest of the

listeners which are likely exhausted by the 15th session. Based on earlier beta-testing and given the context of this training - an advanced ESL pronunciation class which was taken in addition to participants regular graduate work - it was determined that six days of testing and training would suffice.

### 2.9.6 Identification versus discrimination training

L2 speech research has used both discrimination training, in which trainees must say whether pairs of L2 segments are the same or different, and identification training in which the trainees must specify the speech sound they have actually heard. Evidence from research conducted by Jamieson and Morosan (1986, 1989) and Logan and Pruitt (1995) indicates that discrimination training is the less effective than identification training. Strange (1997) asserts that identification training forces the listener to form "some kind of mental representation of the phonetic categories under comparison, instead of directly comparing stimuli on the basis of physical identity alone" (p. 351).

### 2.9.7 Training Vowels

Although the bulk of the non-native phoneme perceptual training literature considers the training of the /l/~/r/ distinction, recent research has begun examining the efficacy of similar training methods for L2 English vowels, (e.g., Iverson & Evans, 2009; Kingston, 2003; Nishi & Kewley-Port 2007, 2008).

Vowel perceptual training comes with its own set of challenges; languages differ in the number of vowels in their inventories, the shape of the vowel space, phonotactic constraints and co-articulation effects. And since co-articulation effects are language specific, training a vowel in

one consonant context may not transfer to other contexts (see Flege, 1995; Strange, 2007; Thomson, Nearey, & Derwing, 2009) and differences in the number of vowels in the inventories between languages can impact the course of non-native vowel learning. For instance, Iverson and Evans (2009) used high variability phoneme training to investigate whether speakers who had a larger L1 vowel space, e.g., German, learned an L2 vowel system (English) more easily than speakers with a smaller vowel inventory, e.g., Spanish. After five training sessions, German speakers improved discrimination at nearly twice the rate of their Spanish counterparts. The Spanish speakers equaled this improvement with an additional ten training sessions. The original hypothesis was that German speakers with an inventory of 18 vowels would have more difficulty with English vowels because of a "crowded vowel space" than Spanish speakers with a five-vowel system.

Kingston (2003) conducted a series of three experiments wherein he trained North American English speakers who had no knowledge of German to identify non-low German vowels. The first experiments showed that some instances of contrastive German vowels were more easily discriminated than others – even when the participants had no previous experience with German. In the second set of experiments, Kingston manipulated the amount of variation by speaker and consonant context in the training stimuli; only those in the paradigm with one speaker but with 18 consonant contexts showed significant improvement in distinguishing German tense vowels. Regardless of variation in the training paradigms, the North American English speaking participants had no difficulty identifying contrasting German high vowels.

Finally, Nishi and Kewley-Port (2007) trained L1-Japanese speakers on English vowels using nonce words under two training paradigms; one included nine North American English vowels and the other focused on the three vowels determined to be most difficult. Learners who

trained on three vowels made rapid progress and achieved high levels of accuracy but made no gains in identifying the other 6 untrained vowels. The features of the "3 most difficult vowels" did not help navigate the rest of the vowel space.

In sum, experience with the L1 shapes the perceptual space of the young learner. The adult L2 learner has difficulty perceiving the subtle acoustic differences of the target language phonology as well as difficulty isolating meaningful from non-meaningful variation in the acoustic signal. Individual sounds of the L2 may be differentially difficult to perceive and acquire depending on their status in the L2 inventory. Coupled with the additional demands of processing language for meaning, phonological learning may be particularly disadvantaged. `

From the training literature it can be surmised that phonological learning requires a context in which a learner can attend to the acoustic or phonetic details that allow for accurate categorization while concomitantly learning to ignore non-significant variation in the signal. Additionally, given the processing load associated with communication—that is, the construction of meaning—training should also limit the amount of semantic and pragmatic processing. Given the general failure of phonological learning "in the wild" and the general success of phonological learning in the lab, extending such methods to the classroom should be a logical progression.

For the work reported on here, training methods from the experimental literature have been repurposed for a classroom context, specifically in conjunction with a graduate-level, ESL pronunciation class. Mandarin and Korean-speaking graduate students participated in vowel training for six tense-lax English vowel contrasts. Training was assigned as homework, over the course of six days, as part of the regular course work.

Chapter 3 reviews the vowel inventories of both Mandarin and Korean based on empirical work that provides acoustic, articulatory and perceptual data. Hypotheses related to the L1 vowel inventory, models of L2 speech perception, and phonotactic constraints are discussed.

**Chapter 3: Vowel Inventories, Phonotactics, and Predictions**

The purpose for reviewing the vowel inventories of the Korean and Mandarin student

learners is to understand what elements of the L1 may impact the perception and learnability of

NAE vowel contrasts, [i]-[ɪ], [æ]-[ɛ] and [u]-[ʊ] that are trained for this thesis.

**3.1 Mandarin**

The vowel inventory of Mandarin is rather variably reported in the relevant literature

with the variation clustering around the number of underlying or phonological vowels versus the

number of "surface" instantiations. Wan and Jaeger (2003) provide an overview of the "6 vowel

system of C. Cheng (1973) which is based on phonemic evidence, the five-vowel systems of R.

Cheng (1966) also based on phonemic evidence; the 5-vowel system proposed by Lin (1989,

1992) based on auto-segmental phonology and under-specification theory, and finally, the four-

vowel system of Wu (1994) based on feature geometry analysis," (p. 208). For this work, I

reference research based on acoustic and articulatory data from Wu (2011) (see also Shih, 1995)

and that of Thomson, Neary, & Derwing (2009).

Thomson et al. (2009) constructed an enhanced discriminate analysis model with a

"modified statistical pattern recognition approach (henceforth referred to as the *metamodel*)" (p.

1449) to measure cross-linguistic similarity between Mandarin and English. The "metamodel"

was trained on vowel categories from both Mandarin and English and then tested on new

productions from each language. Category similarity then was calculated as the number of times

the model categorized an L2 vowel as an L1 vowel and vice a versa. In their model, if two vowel

categories were truly identical, then the metamodel would categorize an instance of a vowel 50%

of the time as English and 50% of the time as Mandarin. Equally, if the two languages have a similar category, the metamodel would be expected to misclassify a vowel of that category at some rate less than 50%.

Wu (2011) in her dissertation examined notions of fluency and accentedness. As part of her research, she evaluated all Mandarin phonetic vowels in all possible syllable combinations using both acoustic and articulatory measurements and compared them to NAE learners of Mandarin. Wu's findings were consistent with that of Thomas (2009); of the vowels trained in this study, [i] was the most similar vowel between the two languages and the most dissimilar were [ɪ, ɛ, u] followed by [æ]. Table 3-1 contains the Mandarin vowels reported in Wu's (2011) dissertation.

Table 3-1 Mandarin vowels (Wu, 2011, p.38)

| Vowel | Count | % of total |
|-------|-------|------------|
| æ | 18916 | 5.17 |
| ɛ | 37070 | 10.12 |
| ɪ | 64388 | 17.58 |
| i | 34042 | 9.30 |
| ʊ | 6194 | 1.69 |
| u | 11584 | 3.16 |

Three of the vowels in Table 3-1 bear the same label as three of the NAE vowels trained in this work, [i],[ɛ][3] and [u].

In Figure 3-1, the mean formant values (from Wu, 2011) of vowels produced in a CV context by eight Mandarin and eight NAE speakers are plotted according to F1 and F2.
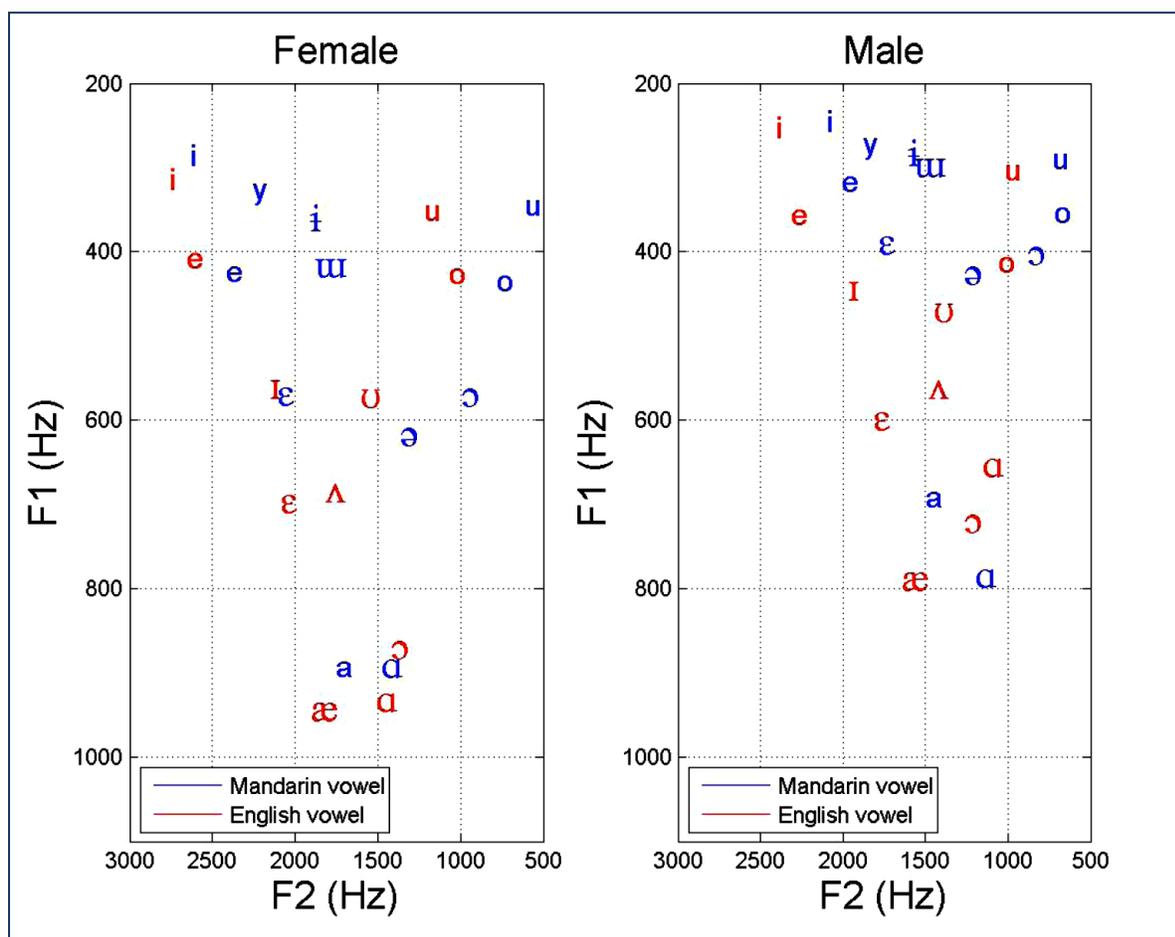


Fig. 3-1 Mandarin and NAE vowels plotted by F1 and F2 from Wu (2011. P. 75)

---

[3] In Mandarin, [ɛ] has a relatively restricted distribution; it only surfaces following the palatal glide [j] and the front, rounded palatal glide [ɥ] in a CGV shaped syllable. Its manifestation then is predictable from its environment. As such, [ɛ] is likely an allophonic variant of /e/. There is a lot of evidence that allophonic variants are more difficult to perceive and learn than sounds that are contrastive (Boomershine, Currie, Hall, Hume, & Johnson, 2008; Strange, 2011; Dupoux, Pallier, Sebastián, & Mehler, 1997; Harnsberger, 2001; Johnson, 2004).

From Figure 3-1, it appears that the differences in formant values between Mandarin and NAE [ɛ] for male speakers are about the same distance apart as that for Mandarin male and female formants for the same vowel. For Mandarin-speaking women, [ɛ] is higher than its NAE counterpart, impinging somewhat on the vowel space of NAE [ɪ]. In addition to phonetic differences with [ɛ], there are also clear differences between the Mandarin and NAE vowel [u] which, for the Mandarin speakers, is further back. Additionally, Mandarin [i] also appears further back and higher, but both Wu (2011) and Thomson et al. (2009) conclude that [i] is relatively similar across the two languages; they also found that NAE [ɪ, ɛ, u] were the least similar to any Mandarin vowel, followed by [æ].

## 3.2 Korean

Korean has 10 vowels, / i e, ɛ, a, ʌ, o, ɨ, u, y, ø/ and a permissible syllable structure consisting of (C)(G)V(C). L1 Korean learners of English have documented difficulty perceiving and producing English vowel contrasts [i]-[ɪ] and [æ]-[ɛ] (Tsukada et al., 2005) and, given that the tense-lax vowel contrast does not exist in Korean, similar difficulties may be assumed for [u]-[ʊ]. Ha (2009) also notes that / ɪ, æ, ʊ/ do not exist in Korean.

Table 3-2 Korean vowels (Ha et al., 2009, p. 20)

| | Front | | Central | Back | |
|---|---|---|---|---|---|
| | Unround | Round | Unround | Unround | Round |
| High | **i** | y | ɨ | | **u** |
| Mid | e | ø | | ʌ | o |
| Low | **ɛ** | | | a | |

Yang (1996) compared ten Korean and 13 American English vowels produced by ten

female and ten male speakers from each language group. In order to reduce variation from

differences in vocal tract length, Yang used a method of uniform scaling within and between the

two languages. This type of transformation "normalizes" the vowel space by increasing or

decreasing distances (between vowels, in this case) by a scale factor that is the same in all

directions; this allows for cross-linguistic comparison of vowels. Significant differences between

the F1 values of NAE and Korean [i] and between F1 and F2 for [u] were reported. Significant

differences in F2 for [ɛ] were reported but only for male speakers. However, the phonemes [ɛ]

and [e] are undergoing a merger among younger Koreans (Igeta & Arai, 2011; Kang, 1996; Lee

& Ramsey, 2000) and the distinction is being lost; as are those related to vowel length. Although

differences between NAE vowels [i][ɪ], [æ][ɛ], [u][ʊ] are qualitative, they also have length

differences and, as such, neutralization of length distinctions in Korean may exacerbate

difficulties with the vowel contrasts presented for training in this work.


**3.3 Plotting the Combined Vowel Space**

Interaction between the L1 and L2 in speech perception and production is complex, but

transfer effects have been well-established in the literature since before Lado (1957) and the

heyday of Contrastive Analysis. It is, therefore, important to examine the relationship between

the formant values for Mandarin, as reported by Wu (2011) and Korean, reported by Yang

(1996) and those of the training materials used in this work. In Figure 3-2, the formant values of

those vowels similarly labeled in Mandarin, Korean, and NAE are plotted F1 by F2. The NAE

formant values were measure at the center of the 3,285 nonce words generated for training, divided by sex of the speaker, and averaged. [4]
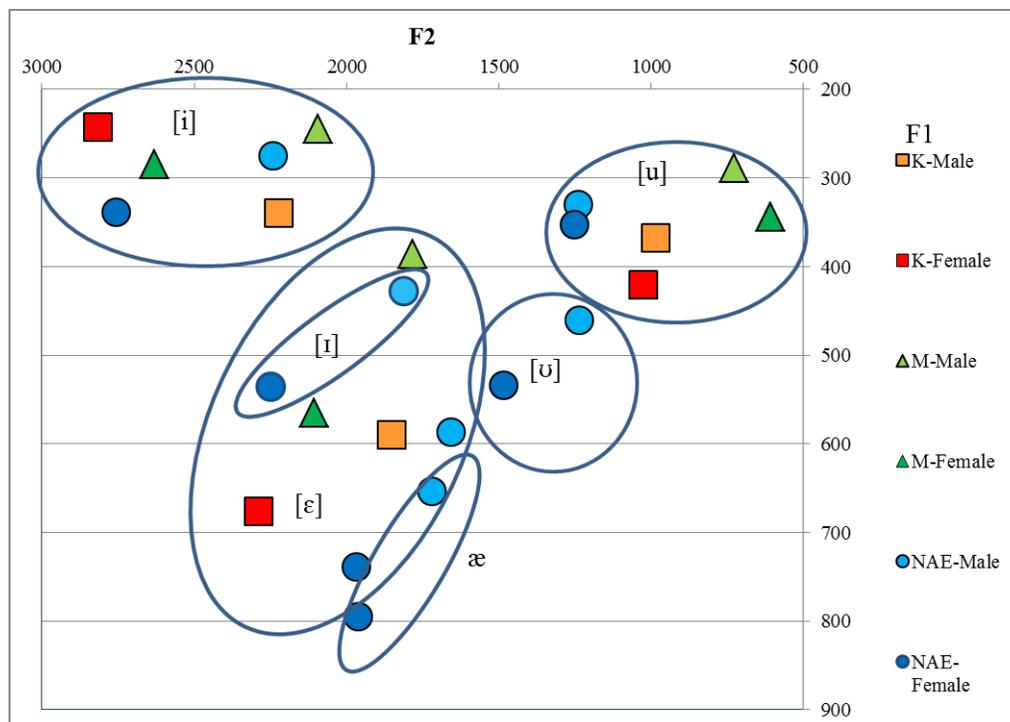


Fig. 3-2 Formant values from Wu (2011), Yang (1996) and from training nonce words

From Fig 3-2, it can be seen that [i] and [u], for all three languages, occupies the most extreme corners of the vowel space. Recall that Yang (1996) found significant differences in F1 between Korean and NAE [i]. Formant values for Mandarin [u], consistent with Wu (2011) and Thomson et al (2009) are higher and further back than the NAE [u] from this data. Yang (1996) found statistical differences in the F1 and F2 values between Korean and NAE [u]. Unlike Wu's

---

[4] Nonce training tokens were CVC with ten varying onsets and codas.

data, NAE [ɛ] and [æ] are not well-separated in the vowel space and appear higher along the F1 dimension. NAE [ɪ] falls within the vowel space of Mandarin [ɛ] and the formant values of male Korean [ɛ] appears to be equidistant from both NAE (male) [ɛ] and [æ].

Restating and summarizing the findings from Wu (2011), Thomas et al. (2009), and Yang (1996):

Table 3-3 Summary of vowel similarity/dissimilarity from Wu, (2011), Thomas et al. (2009) & Yang (1996)

| Wu,Thomson, Yu | æ | ɛ | i | ɪ | u | ʊ |
|---|---|---|---|---|---|---|
| Mandarin | disimilar | disimilar | identicial | disimilar | disimilar | identical |
| Korean | disimilar | ? | similar | disimilar | disimilar | disimilar |

## 3.4 Models of L2 Perception

In the following sections 3.3.1 and 3.3.2, I provided an overview of two models of L2 speech perception, SLM and ASP, respectively that are relevant for the population and training method in this study. Predictions about perceptual acumen in pre-test and/or its development in the course of learning per these models are discussed.

### 3.4.1 SLM

SLM makes predictions about the relative difficulty of acquiring – and producing - a new L2 phoneme in terms of whether the target phone is *identical* to a phone in the L1, *similar* to a phone in the L1 or completely *new* or different from any phone in the L1. *Identical* phones, as one might expect, would be easy to learn; phones that are *similar* are predicted to be difficult to learn since the differences are likely to go unnoticed with the learner associating the L2 sound with the L1 category; *new* phones that are completely different from anything in the L1 should be easy to learn since the differences would be easily discernible. As Flege (1995) puts it, the "greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the

more likely it is that phonetic differences between the sounds will be discerned" and therefore, more readily learned" (p. 239). SLM operationalizes the concepts of *identical, similar* or *new* thus: "L2 vowels represented by phonetic symbols not used to transcribe any L1 vowel have been classified as "new." L2 vowels represented by the same symbol as that used for some L1 vowel have been classified as "identical" or "similar." (Flege, Bohn, & Jang, 1995, p. 440; Flege 1988). This presents difficulties in so far as labeling may depend on the labeler; cross-linguistically, identical labels may be attached to phones that are significantly different phonetically and distributionally. Another difficulty with SLM is that it is unclear how to compute *similar, new* and even *identical* in referring to two sounds – a shortcoming that Flege himself acknowledges given the differences in perceivers, context. and rate of learning.

As such, for this dissertation, two versions of SLM are instantiated; one based on the most common parameters of testing SLM, that is using IPA labels as indicators of *identical, similar* or *new*. As such, [ɪ, æ, ʊ] do not exist in Korean or Mandarin, and as such, they would be considered new sounds and easy to learn. For both Mandarin and Korean speakers NAE [ɛ, i and u] may be considered *similar* sounds since they share the same labels. Given the status of ɛ (as a possible allophonic variant in Mandarin and undergoing merger in Korean) then it would be considered similar but different and therefore hard to hear and learn. NAE [i] and [u] would qualify as identical to Mandarin and Korean [i] and [u]. Table 3-4 summarizes the predictions of a traditional application of SLM.

Table 3-4. Summary of SLM (traditional) predictions.

| SLM - traditional | æ | ε | i | ɪ | u | ʊ |
|---|---|---|---|---|---|---|
| Mandarin | new-easy | similar - hard | identicial-easy | new-easy | identicial-easy | new-easy |
| Korean | new-easy | similar - hard | identicial-easy | new-easy | identicial-easy | new-easy |

Given, however, the availability of formant values, metamodels with confusability matrices, and articulatory information, a second set of predictions are generated using the same labels, *identical, similar* or *new* but using the finding from Thomson, Wu and Yang. These predictions are summarized in Table 3-5.

Table 3-5. Summary of SLM (empirical) predictions.

| Wu,Thomson, Yu | æ | ε | i | ɪ | u | ʊ |
|---|---|---|---|---|---|---|
| Mandarin | new-easy | new-easy | identicial - easy | new-easy | new-easy | identicial - easy |
| Korean | new-easy | ? | similar - hard | new-easy | new-easy | new-easy |

For this case, [ε] has a question mark; if it were just a difference in F2 for male Korean speakers, [ε] would be considered "similar-hard". But given the information that [ε] and [e] are undergoing merger, it's status is unknown.

### 3.4.2 ASP

Strange's (2011) model, Automatic Selective Attention (ASP) would suggest that participants in training conditions such as those provided here are *attentionally focused,* listening for phonetic details, rather than listening in phonological mode. Training stimuli is simple, CVC

nonce words; there are no task demands related to lexical retrieval or meaning construction. As Strange (2011) notes, ". . .when listeners are given simple materials in a perceptual assimilation task. . . their performance reflects their ability to access detailed phonetic information about cross-language and within-language variations" (p. 462). Strange also predicts context-based perceptual differences, similar to those of Flege's SLM.

According to ASP, testing on real words should be disadvantaged since this would constitute listening in "phonological perception mode" which relies on highly automated speech perception routines that tend to ignore variation in phonetic detail in favor of retrieving semantic value. Given that nonce word testing does not present pairs of words to be compared, no predictions can be made. However, in subsequent training, the learners are listening in "phonetic perception mode" which means that they are listening for phonetic information (rather than trying to learn words, for example). As such, according to ASP, the differences between training items should be "perceptually salient" for participants who, therefore, are able to perceive subtle phonetic information that signal differences. They should also then show sensitivity to consonant context. While Mandarin and Korean [i] and [u] may differ from their NAE counterparts, in all three languages they are point vowels, occupying the most extreme areas of the vowel space. As such, despite variation between languages and how these same vowels might vary as a result of co-articulation effects, they should be relatively easy for both L1 groups to identify. NAE [ɪ] is relatively close to Mandarin [ɛ] and as such, Mandarin speakers may have an advantaged at the outset identifying [ɪ] (vs. [i]). Korean participants in the pretest may have difficulty identifying [ɪ] but in training, while in "phonetic perception mode," they should quickly become attuned to the difference. Additionally, for both groups, given the relatively good category separation of [i] and [ɪ],in the training material (Fig. 3-2), improved accuracy should be clearly evidenced by the

end of training. If Korean and Mandarin speakers are good at identifying [u], then in training it should be easy to disambiguate [ʊ] even if in phonological perception mode both groups have difficulty perceiving the difference with [u]. Given good category separation in the training material and the sensitivity to phonetic detail predicted by ASP, identification and disambiguation of [u] and [ʊ] should progress rapidly.

Korean speakers should have some advantage with [ɛ] and [æ], at least initially. If NAE and Korean [ɛ] are phonetically similar and NAE [æ] is quite distinct from Korean [ɛ] then accuracy in identifying this distinction may be relatively easy. However, NAE [ɛ] and [æ] are relatively close in the vowel space; as Korean learners go through the training and learn the subtle differences between Korean and NAE [ɛ], the category boundary for NAE could shift. As such, [ɛ] and [æ] will be much closer together which, together with novel coda effects, may serve to blur the distinction rather than sharpen a category boundary.

For Mandarin speakers, the status of [ɛ] as an allophonic variant should make it particularly difficult to initially perceive and to learn (Strange 2011). This may be compounded by the poor separation of [ae] and [ɛ] in the training material.

Table 3-6 provides a summary for ASP predictions by L1 and vowel separated into "phonological perception mode" and "phonetic perception mode".

Table 3-6. Summary of SLM predictions by L1 and vowel.

| ASP - Mandarin | ae | eh | iː | ɪ | u | ʊ |
|---|---|---|---|---|---|---|
| phonological perception | hard | hard | easy | hard | easy | hard |
| phonetic perception | easy | ? | easy | easy | easy | easy |
| | | | | | | |
| ASP - Korean | ae | eh | iː | ɪ | u | ʊ |
| phonological perception | hard | easy | easy | hard | easy | hard |
| phonetic perception | easy | easy | easy | easy | easy | easy |

## 3.5 Phonotactics and Coda Conditions

While the L1 vowel space is important for hypothesizing about the relative ease or difficultly in acquiring non-native vowels, the effects of coda constraints (Halle, Best, & Backrack, 2003; Strange, Akahane-Yamada, & Kubo, 2001) and co-articulation effects can equally contribute to perceptual difficulties. Although both Strange (2011) and Flege (1995) predict that consonant context may impact the identification of non-native phones, parameters that might allow one to predict what and how context are not provided. I default here to the assumption that what is absence in the L1 will be difficult in the L2.

The consonant context is predicted to impact identification of vowel contrasts. Mandarin has a relatively restricted syllable structure: C(G)V(G) and C(G)V(N) [5] where (G) and (N) are optional and VGN is not a licit syllable. If there is a consonant coda, it is either [n] and [ŋ]. Given the array of co-articulation effects associated with substantially more codas (in the training material), Mandarin speakers may have initial difficulty except with nasal codas.

Korean has a much broader range of licit codas, /p/, /t/, /k/, /m/, /n/, /ŋ/, and /l/ (Ha, Johnson, & Kuehn, 2009) five of which overlap with the codas in the training nonce words, /p, t, k, m, n/. While Korean does not have voiced stops in its consonant inventory, it does have an intervocalic voicing rule, (Silva, 1992; Jun, 1996). For example,

papi "rice"        ibabi "this rice (nom.)"

kuk "soup"        igugi "this soup (nom.)"

tʃip "house"        idʒibi "this house (nom.)"

---

[5] C is a consonant, G, a glide, V, a vowel and N is a nasal.

As such, those vowels with /b, d, g/ in coda position are predicted to present no more difficulty in terms of vowel identification, than those with /p, t, k/ since co-articulation effects between vowels and obstruents with +/- voicing are encountered in Korean. The only coda in the training material that has no counterpart is /s/ and is predicted to provide another dimension of difficulty in vowel identification.

**3.6 Variability in Training Material**

Another area of investigation is related to the amount of variation present in each of four training paradigms. It is now prevalent in the literature that high variability training, provides the most robust category learning. High variability training most often refers to multiple speakers. Although "speaking rate" is mentioned in a few training studies, it is rarely controlled for as a training variable. If we assume that multiple speakers provide an array of phonetically varied exemplars of a category, then part of the phonetic variation includes co-articulation effects. Equally, the faster the speech rate, the greater the co-articulation effects. If learning a new vowel contrast is also learning how it varies then rate changes provides one kind of information and multiple speakers provides another.

As mentioned, it is prevalent in the literature that high variability is the most effective; there is, however, some evidence that certain, highly confusable sounds may benefit from low-variability training that repeatedly presents something more akin to a category prototype (e.g., Jongman & Wade, 2007; Perrachione et al., 2011; Shih et al., 2010,). Finally, while categories may be more or less confusable, differences in individual acumen may also interact with the amount of variation per paradigm. If a learner starts off with relatively poor perceptual discrimination, then low variability may be more helpful, particularly in the initial stages of learning.

Chapter 4, Methods, provides an overview of the design and implementation of the training protocols including the generation of the training tokens, recording procedures, rational for the use of nonce words, pre training familiarization and details relating to the profiles of the participants.

## Chapter 4: Methodology

### 4.1 Overview

The perceptual training of North American English (NAE) tense-lax vowel contrasts reported on here was adapted from the perceptual training literature but carried out in the context of a graduate-level ESL pronunciation class. Participants were matriculated graduate students poised to become Teaching Assistants (TAs) in their respective programs. The majority of the students had taken and failed a university-internal oral proficiency interview and were required to take a pronunciation class, (or complete 10 hours of tutoring), prior to retaking the exam. The vowel training was assigned as part of standard course work although students had the option to not allow their results to be used for research purposes. The delivery mechanism for the training was a web-based application that could be accessed from any computer with an internet connection. The application provided a simple user interface in a format with which most people are familiar, e.g. "click and play" with radio buttons to indicate selections.

Prior to training, student participants completed a pre-test on 39 real word minimal pairs. Real word minimal pairs were presented to students on a screen. Once the page opened, the sound file played automatically, one time. Student participants had the option of listening one additional time but were then forced to make a selection. Half of the real words presented to the students were the higher frequency member of the pair while the other half was the lower frequency member. Each vowel pair was presented 13 times (three vowel pairs x 13 = 39).

After completing the real word pre-test, student-participants were familiarized with the spelling conventions that were adopted for the nonce training items. Familiarization took place in the classroom and included both visual and aural presentation. They were then assigned the pre-test on nonce words as homework.

Students were randomly assigned to one of four training paradigms (detailed below) and then instructed to complete a pre-test on nonce words as homework. Students had to be registered to participate in the training and were required to log in to access each day's training and testing. Timestamps were generated at login and keystrokes indicating which sound files were played and how many times (in the training session) were recorded. When the student chose to finish the training session, he/she was then tested on a subset of that day's training material. Over the course of six days students logged in and listened to as many contrasting minimal pairs as they wanted, and then completed a daily post-test on a subset of that day's nonce-words.

For two of the training paradigms, Paradigm A and B, the nonce minimal pairs were produced by eight speakers of general North American English, balanced for sex; training Paradigms C and D consisted of the same nonce minimal pairs, presented in the same order but produced by only one speaker. In Paradigms A and C, the speakers' rate increased on the third and fifth day. The presentation of stimuli was subject-controlled, thus allowing subjects to compare both targeted vowels and speakers voices. In the daily testing phase, participants received immediate feedback, and in the case where an incorrect choice was made they had the opportunity to listen again. The following section provides details on the generation of the training stimuli, recording procedures, construction and parameters of the four training paradigms.

**4.2. The Training Stimuli**

**4.2.1 Nonce minimal pairs.**

Single-syllable, nonce minimal pairs of the form CVC were generated such that ten consonants (p, t, k, b, d, g, s, m, n), controlled for voicing, place, and manner-of-articulation, were distributed between onsets and codas of three tense-lax vowel pairs [æ~ɛ], [iː~ɪ] and [oʊ~ʊ]. The consonant [h] was restricted to onsets, and [s] was restricted to codas. A total of 486 tokens resulted from the combination of nine onsets, six vowels, and nine codas. There is some debate in the literature regarding vowel identification. One view is that the formant structure of the steady-state vowel provides the relevant spectral cues to vowel identification, and that these cues are particularly robust in noise. The other view is that the spectral changes at syllable edges contribute more to vowel identifiability. The latter position is supported by experiments in which the vowel portion of a CVC utterance was extracted or replaced by noise. Listeners were more accurate at identifying vowels in silent center CVC contexts than at identifying steady state vowels presented without context cues (Jenkins et al., 1983; Strange et al., 1983, 1976). These findings were further supported by Nittrouer (2007) who compared adults and children on their perception of vowels and found that the dynamic structure of the transitional features were most important to vowel perception. Regardless, co-articulation effects due to consonant environment are an important element of variation and hence, are an important element of training.

**4.2.2 Rationale for using nonce words.**

The rationale for using nonce words was multifold. Nonce words allow for control of the idiosyncratic nature of English spelling. As noted by Logan and Pruitt (1995), "A significant problem with the use of the identification task in cross language research is that labels for the

non-native phonemes must be provided to the listeners. What constitutes the appropriate label is not obvious" (p. 359). In the case of NAE vowel training, given the orthographic system of English with its variable spelling of sounds in the same category, (e.g., *foot, would, put* or *boot, rude, lewd*), a consistent one-symbol-to-one-sound convention had to be instantiated. In additional to consistent labeling, it was also necessary to consider the accuracy of the mapping between a previously learned orthographic representation and a matched phonological representation.

Orthography interacts with phonology in lexical representations in several ways. Schiller (1998) attests to "lexical-phonological priming" such that the orthographic representation of a familiar word "primes" the inaccurate phonological form stored in the learner's lexicon. In their research training with listeners with cochlear implants, Li and Fu (2007) note that a lexical item encountered and stored (along with its label) during 'normal hearing' can interfere with training for new, spectrally shifted, acoustic patterns. The label of the previously stored lexical item creates a bias such that the learner fails to recognize the new, spectrally shifted lexical item.

More recent research has provided evidence that knowledge of an orthographic form can influence the learner's memory for the phonological forms of L2 words, and that orthographic representations can facilitate or inhibit the encoding of new lexical items, (Bassetti, 2006; Escudero & Wanrooij, 2010; Escudero, Hayes-Harb, & Mitterer, 2008; Hayes-Harb, Nicol, & Barker, 2010; Simon, Chambless, & Alves, 2010; Showalter & Hayes-Harb, 2013). Given that much L2 learning happens in the classroom, it is likely that a significant amount of the input learners receive is written rather than spoken (Bassetti, 2006). As such, it's likely that orthographic representations uncoupled to accurate phonological representations will impact both production and perception in ways that are just beginning to be researched.

In addition to confounds due to orthographic representations associated with inaccurate phonological forms, nonce words were chosen for training to avoid frequency effects. For example, in the Buckeye Speech Corpus (Pitt et al., 2007), 36, 6203 segments were labeled as vowels. Of these, [ʊ] occurred 6,194 times, which represents 1.7% of the total vowel inventory of the corpus. In contrast, [ɪ] occurred 64,388 times, representing 18% of the total. Relevant vowel frequency counts from the Buckeye spontaneous speech corpus are reported in Table 3.1 (Note that vowel labels are in IPA). The Buckeye corpus was selected because of the detail of the phonetic transcription.

Table 4.1 Vowel counts from the Buckeye Speech Corpus

| Vowel | Count | % of total |
|:-----:|:-----:|:----------:|
| æ | 18916 | 5.17 |
| ɛ | 37070 | 10.12 |
| ɪ | 64388 | 17.58 |
| i | 34042 | 9.30 |
| ʊ | 6194 | 1.69 |
| u | 11584 | 3.16 |

Lexical items that occur more often in the target language can bias the participants' performance on perceptual training tasks. In other words, if a student had encountered a common word in English, such as *read,* multiple times, their familiarity with the word would be expected to be higher than for a low frequency word, for example *rid*. Multiple encounters with the same

word increase the likelihood that a participant would be able to correctly identify it, circumventing phonemic level training. Flege et al. (1996), found this to be the case in their 1996 study, where Experienced Japanese (EJ) learners and Inexperienced Japanese (IJ) learners were better able to correctly identify /l/ and /r/ tokens in more frequent words than those in less frequent words. Additionally, this same familiarity with vocabulary might allow "top down processing" so that students might employ other cognitive skills to determine the phoneme. MacKay (1987) notes that information about the lexical item can overrule acoustic information such that a target word is identified and processed before it is even physically heard. In this way, the hearer fills in missing information, deciding what they have heard before they have actually heard it.

Finally, using nonce words provides a better test of the validity of exemplar models of encoding speech. Since there is no way to know how many times an individual has previously encountered any given lexical item, (outside of estimates provided by frequency counts), nor any way of accessing the phonological representation that the listener already has stored, training with nonce words provides a *tabula rasa* for novel exemplar encoding.

**4.3 The Talkers and Recording Procedures.**

The speech materials used in the perceptual training experiments reported here were produced from ten speakers of North American English, and were recorded in the phonetics lab at the University of Illinois. The speakers consisted of five men and five women ranging in age from 22 to 28 years whom were speakers of standard North American English (NAE). Eight of the speakers' recordings were used for training; the ninth and tenth speaker provided pre-test and post-test materials for both real and nonce words and served as a "novel speaker/voice" in order

to check for the learners' ability to generalize to novel exemplars. The speakers were digitally recorded in the phonetics lab on the University of Illinois campus using a Shure SM7A microphone set for flat dynamics. The speech signal was amplified through a Mackie 1402-VLZ Pro mixer with a flat equalizer (EQ). That signal was digitized with an M-Audio FireWire 410 A/D Interface at 44.1 kHz and 16-bit quantization. Speakers were asked to produce a carrier phrase with the targeted nonce lexical item "I say *<nonce>*, now" as a complete intonational phrase. Superfluous prosodic features, such as phrase-final lengthening, list intonation contours or contrastive pitch accents were controlled for in the sense that recording was interrupted and speakers asked to repeat if something other than a complete intonational phrase with a falling pitch accent was produced.

For the first recording, speakers were asked to produce the training stimuli in a clear and careful manner, as though they were demonstrating pronunciation using (slightly) exaggerated speech. For the second set of perceptual stimuli, speakers were asked to repeat the same nonce tokens in a "normal" or casual style as though they were speaking to an L1 NAE speaker whom they did not know. The third and final rendition was at a fast speech rate such that speakers were asked to produce the nonce pairs quickly, as though they were speaking in a very causal situation with a friend. Under each instructional condition, speakers repeated the tokens in carrier phrases three times.

A total of 3,285 nonce words were recorded. The target words were extracted from the carrier sentence using Praat (Boersma, 2001) and then binned according to six dimensions: (1) sex of the speaker, (2) speaker identity, (3) tense-lax vowel pair, (4) target vowel, (5) consonant context in both onset and coda, and (6) rate of speech, (careful, casual, fast).

## 4.4 The Training Paradigms

The training paradigms were created so that the presentation order of all the training words was exactly the same. Training Paradigm A was created using 600 unique training tokens such that 100 tokens were presented each day. For the first two days of training, a balanced set of tokens was selected only from the "careful" productions of all eight speakers. For the two subsequent days of training, a balanced set of speech tokens were selected from the "casual" pool and the last two days of training from the "fast" tokens. Paradigm B mirrored training Paradigm A with the same list of training words, presented in exactly the same order, and produced by the same speakers, but with only nonce words selected from the "careful" bin presented for training. Participants in Paradigm C were again presented the same words in the same order as A and B, and like "A" the words increased in speed over the course of training but were tokens from only one speaker. Similar to C, Paradigm D presented the same words, in the same order, produced by one speaker, but all were from the "careful" bin. After parceling out 'bad' sound files, students in Paradigm A trained on 576 unique nonce words; Paradigm B, 534 unique nonce words, C was 438 and D 252 unique nonce words[6]. Each day of training was followed by a test of 25 words from the training material of that day. Table 4-2 summarizes the characteristics of the training paradigms:

---

[6] Note: each talker repeated each nonce word in a carrier phrase three times at three speech rates. While the listener may have encountered the same talker producing the same nonce word, it was not the exact same sound file.

Table 4-2 Summary of training paradigms

| Training A | Multiple speakers, Multiple speech rates | 576 nonce words |
|------------|------------------------------------------|-----------------|
| Training B | Multiple speakers, Single speech rates   | 534 nonce words |
| Training C | Single speaker, Multiple speech rates    | 438 nonce words |
| Training D | Single speaker, Single speech rates      | 252 nonce words |

## 4.5 Departures from Other Training Protocols

Note that the training paradigm developed for this work differs from typical High Variability Phoneme Training along several dimensions. Much of the HVPT literature employs multiple talkers but not within one session. Most often, a unique talker provides all the training material for one training session, with speech samples from different talkers presented in different training sessions (e.g., Lively et al., 1993; Lively et al., 1994; Logan et al.,1991). In the training protocols designed for this study, those in the multi-talker paradigms, A and B, above, exposed listeners to eight different speakers on each day of training.

Additionally, as noted in Wanrooij and Boersma (2013), most training for non-native category learning has relied on bimodal distributions of stimuli with few different values or tokens; e.g. training on one minimal pair *lock~rock* with more or less exaggeration of the target phoneme. In this work, 486 naturally produced tokens constitute the training material. And again, as Wanrooij and Boersma, (2013), note repeated natural tokens are rarely identical and as such provided richer sources for learning.

Perception studies have found a disruptive effect for recall of word list when there were multiple speakers or different rates of speech (differences in volume had no effect.). While

multiple talkers as a training condition has been empirically testing in the perceptual training literature, speech rate has not, to my knowledge, been tested empirically for any effect on training.

Finally, as noted earlier, most perceptual experiments and many of the vowel training work, has relied on CV structures. In this work, nonce words were generated with nine different codas in order to explore the effects of co-articulation on L2 vowel perception and learning.

## 4.6 The Sound Files

The nonce word sound files were run through a forced aligner (P2FA, Yuan & Liberman, 2008), text grids were automatically generated with labeled intervals, then hand corrected; duration and formant values were extracted from the intervals labeled for each of the vowels. F1-F2 by F1 values is plotted in figures 4.1 through 4.4, which provide a representation of the amount of variation and category overlap in each training paradigm.
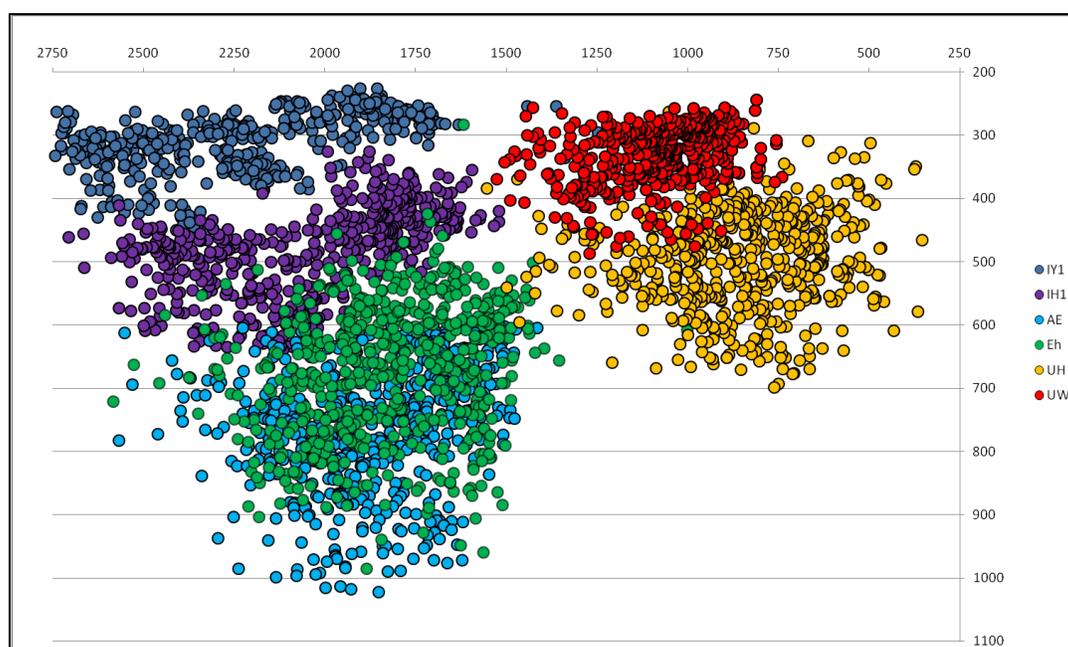


Fig. 4.1 All vowels from all the training token
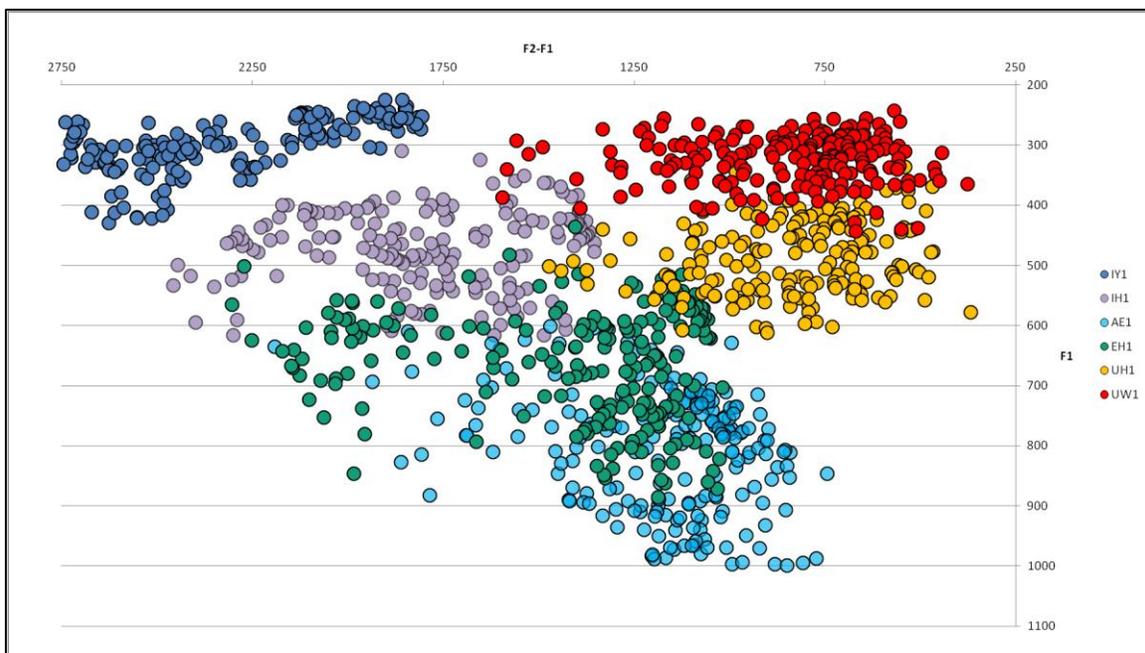
Fig. 4.2 Paradigm A—Eight Speakers, Three Rates
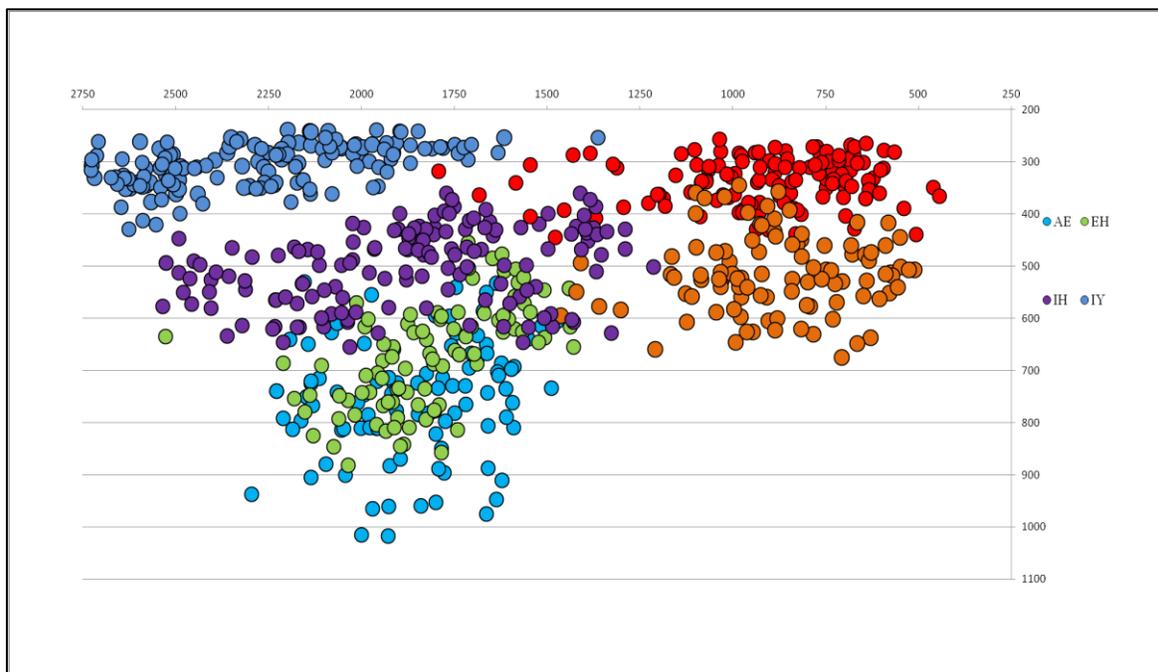


Fig. 4.3 Paradigm B—Eight Speakers, One Rate
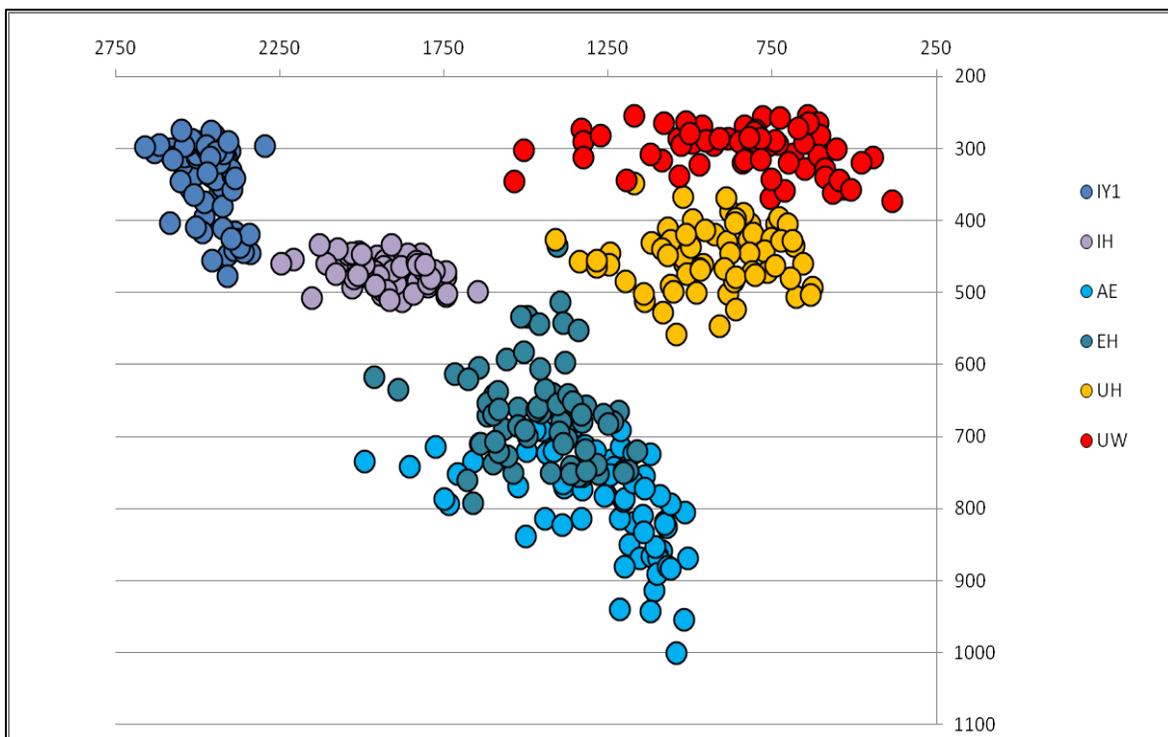
Fig. 4.4 Paradigm C—One Speaker, Three Rates
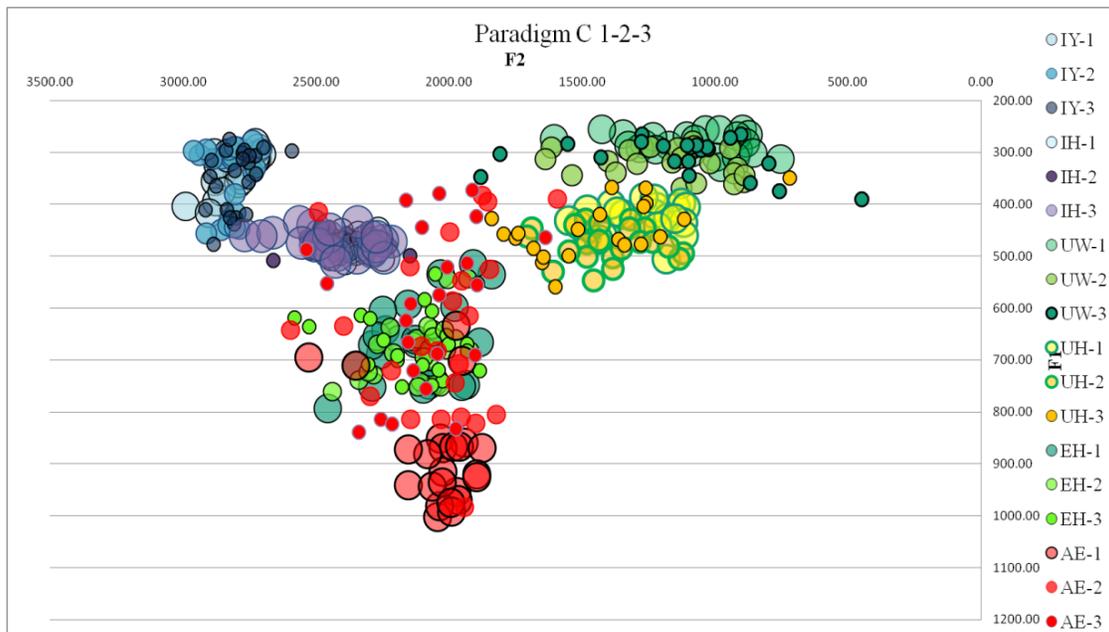


Fig. 4.4 Paradigm C—One Speaker, Three Rates (plotted with rate differences
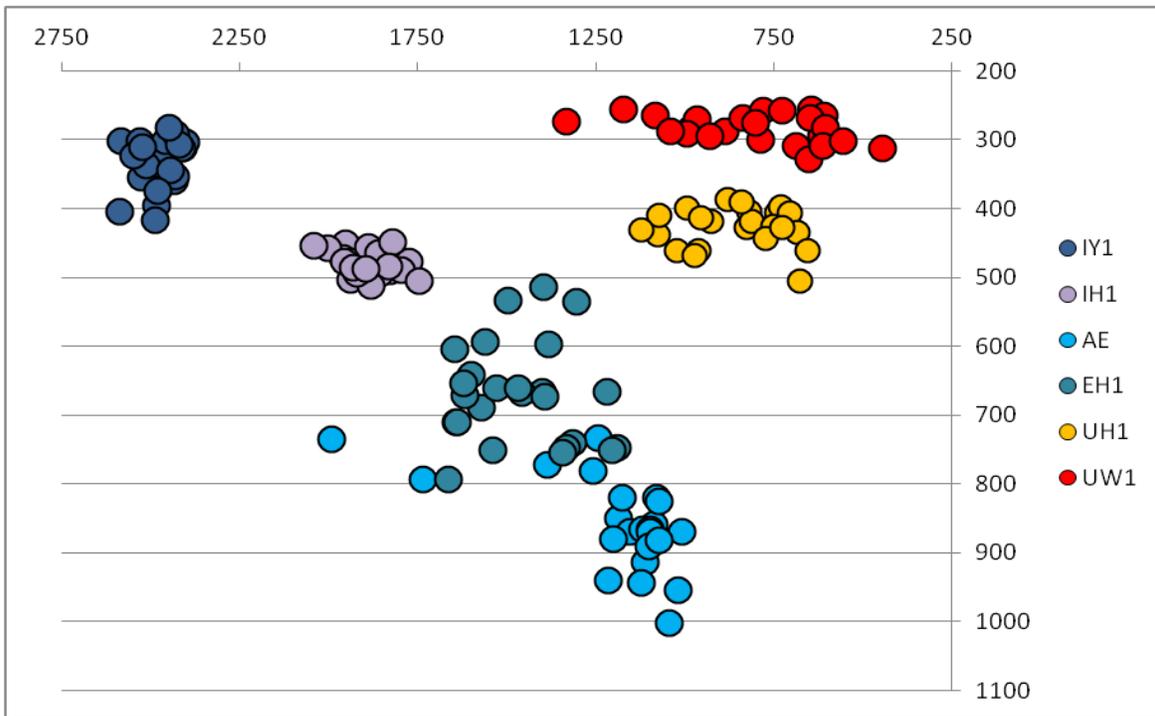
Fig. 4.5 Paradigm D—One Speaker, One Rate

The nonce tokens were produced under three speech styles: careful, casual but clear, and very casual and fast. The differences in speech rate are represented in terms of the duration of each vowel at each of the three rates in Fig. 4.6:
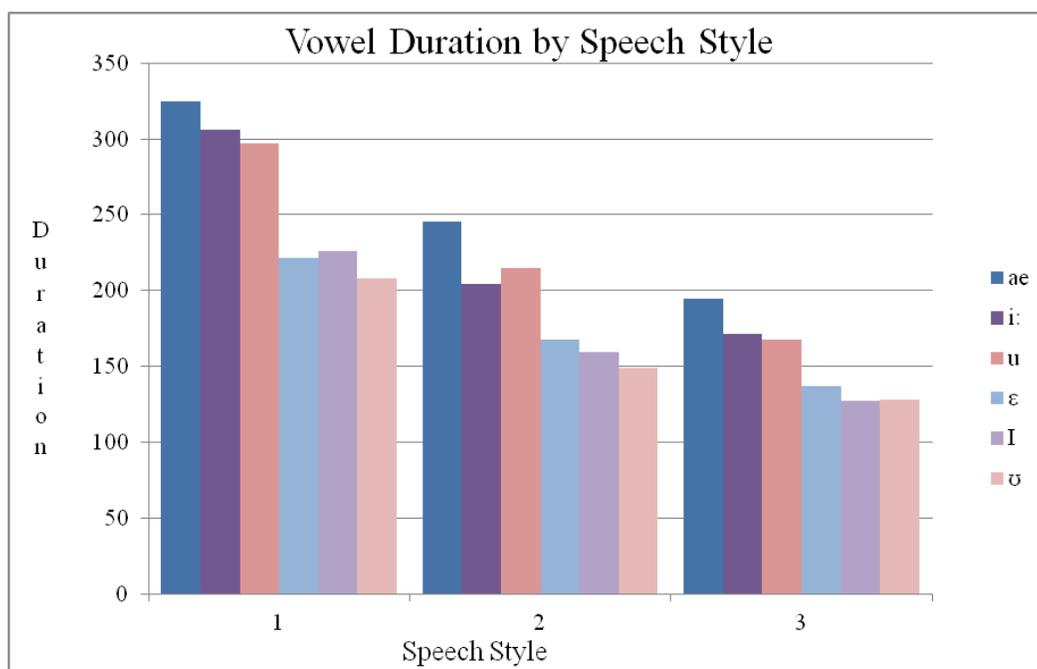
Fig. 4.6 Vowel Duration by Paradigm

The training material was validated, serially, by four L1 speakers of NAE; any tokens that the first L1 listener deemed to be off target were discarded and replaced with a token from the pool of remaining tokens. Then the whole set of materials were listened to and assessed again by a second L1 speaker. This process was repeated four times.

## 4.7 Procedures

Prior to any training, all participants completed a pre-test on 39 real words, balanced for frequency. Half of the test items presented for identification were the higher frequency member of the minimal pair and the other half were the low frequency item. The testing consisted of two orthographic representations of real words presented on each screen. Upon opening the screen, there was a one second delay and then the test item played automatically. There was a "play"

button so that students could re-play any test item one additional time but no more, and then had

to select one of the two options. Students received immediate feedback on their selection. The

same real-word test was administered at the end of training.

As noted above, the training material consisted of nonce words, which allowed some

control over the quixotic nature of English spelling and provided a consistent one-symbol-to-

one-sound convention. However, participants had to be trained on the labeling convention prior

to perceptual training. The labeling/spelling conventions were as follows:

Table 4-3 Trained Spelling conventions

| Spelling convention | IPA |
|---|---|
| ee | /iː/ |
| i | [ɪ] |
| ae | [æ] |
| e | [ɛ] |
| oo | /uː/ |
| ou | [ʊ] |

All participants received in-class training on the spelling conventions. After providing the

symbol-sound match and giving examples, students were trained in the following manner. Nonce

words, which were similar to those in the vowel training application but with a greater variety of

onsets and codas, were presented on a screen. Students repeated each nonce word after the

instructor, going through a set of 105 slides relatively carefully. In the second repetition, students

said the nonce words all together. If the instructor heard an off-target production, students

repeated the item. On the third repetition, the same nonce words were flashed on a screen in

increasingly rapid succession at which point the student-participants had to say the nonce word

as quickly as possible. All participants went through 105 slides twice, as a group. Each student

then repeated the naming task individually, with correction as it was warranted. All students

were then instructed to complete a pre-test on nonce words at home that evening.

The student participants, excluding the control group, were instructed to log in each day

for six days and complete a session of training and testing. In a departure from the training

models discussed in sect. 2.8, training and testing were constructed as separate activities. In the

laboratory-based training literature, training typically consists of a series of forced-choice tasks

with immediate feedback; the training is, in effect, also testing. In at least one study by Golestani

et al. (2004), they used functional imaging to compare brain activation of "good" vs. "poor"

learners during novel phoneme training. They found that some of the poor learners reported

strategies in which they compared each sound to the previous one, and that "their performance

tended to deteriorate every time that they received negative feedback on an item" (p. 503). In

order to avoid students implementing similar strategies, in the design employed here, training

consisted of 5 minimal pairs, 10 tokens, per page. Students could play and compare tense~lax

pairs, e.g. [iː] and [ɪ] or they could, on the same page, play the [ɛ] from the [ɛ]~[æ] minimal pair

and compare it with the [ɪ]. In the training component, there were no right or wrong answers, just

100 sound files consisting of tense~lax vowel pairs. Students could spend as much time as they

wished or felt they needed and were free to move between and among the nonce minimal pairs.

Testing took place when the student chose to end training. The test that followed each day's

training consisted of 25 words, which represented a subset of that day's training material. Figure 4-7 is a screen capture of a training page.
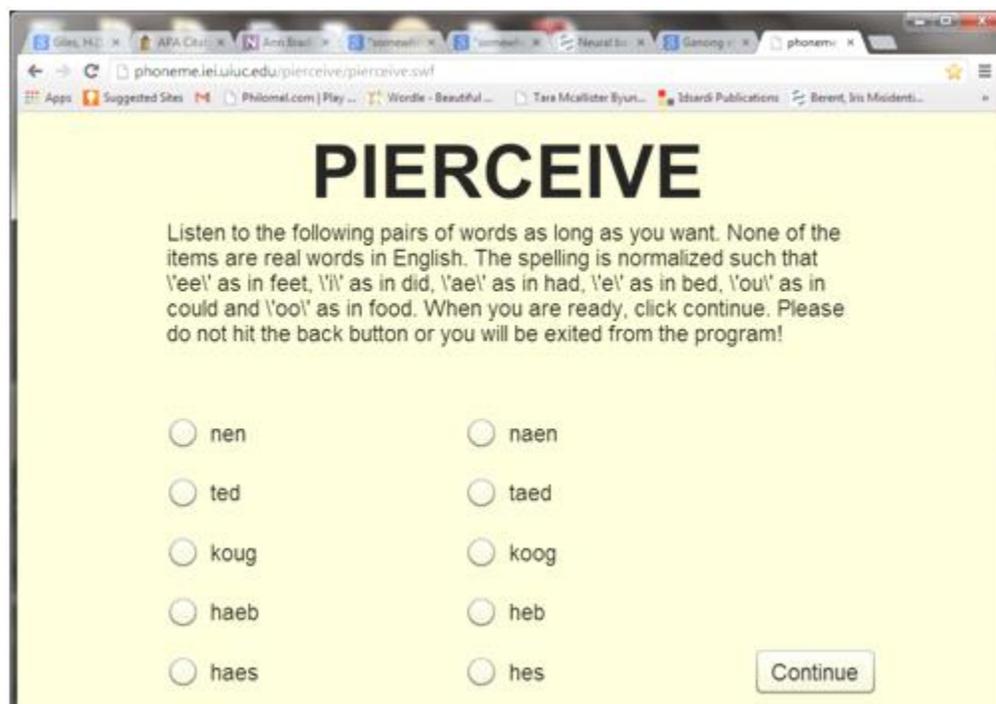


Fig. 4-7 Screen capture of training application

It should be noted that the pronunciation class in which the students were enrolled met twice a week; as with any homework or assigned task, they had control over when they logged in and how much time they spent on task. That being said, the application was designed to track the behavior of the learner through the course of their training and testing. A time-stamp was generated when a student logged in and when they competed the session. Each time a student listened to a nonce word, the keystroke was recorded. The testing section was designed such that listeners could hear a test item twice. They then had to select the text corresponding to the sound file of the test word. They could change their selection (which was also recorded) as many times

as they wanted before hitting a "submit" button, which recorded their final choice. In this way, there was a record of which phonemes students played, how often, in what order, and in which pairs, (however, analysis of serial order and pair-wise order is left for future work).

After completing six days of training and testing, participants were given a forced-choice post-test on the same nonce words encountered in the pre-test. The following day, students completed a post-test on real words, the same words that they had encountered in the pre-test, produced by the same speaker, although they had not heard that particular voice anywhere else in the course of training. The control group only received the in-class training on spelling conventions and completed the pre-test and post-test for real and nonce words. The pronunciation course lasts a full semester, 16 weeks. The pretests, training, and posts tests were completed within the first three weeks of the course. The flow chart in Figure 4-8 provides a summary of the procedures.

Fig. 4-8 Flow chart summarizing procedures

| Pre-Test on **Real Word** | → | Training on **Spelling conventions**: "Lightening | → | Pre-Test on **Nonce Word** Minimal Pairs |

| Prdgm. **A:** 8 voices | Prdgm**. B:** 8 voices | Prdgm. **C**: 1 voice | Prdgm. **D**: 1 voice |
| **1.Train** 50 minimal | **1.Train** 50 minimal | **1.Train** 50 minimal | **1.Train** 50 minimal |
| Test 25 | Test 25 | Test 25 | Test 25 |
| **2.Train** Test **3.Train** Test **4.Train** | **2.Train** Test **3.Train** Test **4.Train** | **2.Train** Test **3.Train** Test **4.Train** | **2.Train** Test **3.Train** Test **4.Train** |

Post-Test on **Nonce Word** Minimal Pairs

Post-Test on **Real Word**

At the end of the course, students were given an online real-word familiarity survey on the 78 real words presented in the pre- and post-tests. Student participants had the following choices for each word: 1. *Very familiar - it's a common word that I hear and use all the time*; 2. *Somewhat familiar - I know this word but I don't really use it*; 3. *I'm not sure - I \*might\* know this word from a vocabulary test*; 4. *I've never seen this word before*! Of the 39 words on the real word pre-and post-test, 17 were identified as "*Very familiar - it's a common word that I hear and use all the time*" – response number (1) by 93% of participants and 19 of the words were identified as either response 3, *I'm not sure - I \*might\* know this word from a vocabulary test*; or 4 above, *I've never seen this word before!* Three of the items were distributed nearly equally along the scale but in general, roughly half of the items were very familiar to students and the other half relatively unknown.

## 4.8 Participants

One of the more difficult aspects of L2 research is the number of variables that can affect L2 learners' developmental trajectories. For this research, the variation among these factors was limited, albeit serendipitously. All the subjects were students admitted to the graduate college at the University of Illinois. As such, the graduate college requires a minimum score of 79 on the Internet-based Test of English as a Foreign Language (TOEFL iBT). On the speaking section the

TOEFL iBT, participants scored an average of 20.38 (out of 30) with a median of 20 and mode of 22[7].

The majority had been offered teaching assistantships but had not passed an oral proficiency interview. All were enrolled in an advanced pronunciation class. Participants were 61 Mandarin and 41 Korean L2 English Speakers; seven Korean and eight Mandarin speakers served as control, but were offered the training at a later date.

Participants ranged in age from 23-34, with a mean age of 25.9. The average length of residence in the U.S. was 2.5 years, but ranged from six months to eight years. All had begun formal English instruction in primary school in their home countries.

The next section, Chapter 5, reports on the results of the pre-test, training and post-tests for the Mandarin and Korean graduate students who participated in this study.

---

[7] A score of 24 or above on the speaking section of the TOEFL iBT is sufficient for an international graduate student to teach without restrictions and is not required to participate in any additional ESL speaking/pronunciation courses.

**Chapter 5: Results**

In order to assess the effects of the training paradigms, there are several variables whose relative contributions to any improvements in vowel perception need to be examined. Improvement is defined as percent change on post-test relative to pretest; scores were computed separately for real words, nonce words and for individual vowels.

**5.1 Statistical Model**

Improvement in vowel perception is analyzed in relation to a number of factors using linear regression. The predictor and response variables are described further here. There are two populations of L2 learners, Mandarin (N=53) and Korean (N=34) graduate students. Mandarin Chinese and Korean have different vowel inventories and different coda constraints which may impact the effects of training; so language group (L1) is a categorical predictor variable. Individuals from each L1 group were randomly assigned to one of four training paradigms that varied by the elements of the training; the second categorical predictor variable is Training Paradigm. Within each of the training paradigms, learners controlled the number and type of training tokens they listened to. The number of tokens played in the course of training can be understood in the amount of time on task – which may also impact outcomes. Total practice time is the continuous predictor variable. Therefore, the predictor variables for a generalized linear model, are L1, training paradigm, amount of time spent on training and the baseline condition of participants as determined by pretest scores. The response variable is the percent change between pretest and post-test scores. The general linear model then, is Percent Improvement predicted by L1, Training Paradigm, Pretest score and/or total training. Results (Table 5-1) indicate that the

Training Paradigm (A, B C, or D) and Pretest score for nonce words are the most significant

predictor variables.

Table 5-1 Results of the general linear model: Regression statistics

| | | | | | |
|---|---|---|---|---|---|
| Multiple R | 0.66918 | | | | |
| R Square | 0.4478 | | | | |
| Adjusted R Square | 0.41371 | | | | |
| Standard Error | 7.66648 | | | | |
| Observations | 87 | | | | |
| | *df* | *SS* | *MS* | *F* | *Signif. F* |
| Regression | 5 | 3860.69 | 772.138 | 13.1372 | **< 0.000** |
| Residual | 81 | 4760.77 | 58.7749 | | |
| Total | 86 | 8621.46 | | | |
| | *Coefficient* | *StndrdEr* | *t Stat* | *P-value* | |
| Intercept | 54.182 | 8.43599 | 6.42272 | **< 0.0001** | |
| L1 | -1.7299 | 1.85339 | -0.9334 | 0.35339 | |
| RealWord Pre | 0.24987 | 0.20476 | 1.22032 | 0.22588 | |
| Nonce Pretest | -1.989 | 0.27291 | -7.2881 | **< 0.0001** | |
| Paradigm | -2.5919 | 0.75325 | -3.441 | **< 0.0001** | |
| Total Training | 0.00025 | 0.00092 | 0.27239 | 0.78602 | |

Since the Pretest score on nonce words and the Paradigm in which the student trained are

significant predictors of percent improvement, differences among means is explored in a series

of post-hoc comparisons.

**5.2 Pretest on Real Words**

The student-participant groups were relatively homogenous in terms of formal education,

TOEFL scores, age at which they started formal English instruction and length of residence in

the US. Nonetheless, in order to establish some baseline measurement of perceptual acumen

prior to training, students completed a pretest on 39 real-word, minimal pairs, half of which were

the higher frequency member of a minimal pair and the other half of which were the low

frequency member of the minimal pair. Frequency counts were taken from the combined

language frequency measures at Brigham Young's Corpus of Contemporary American English

(http://corpus.byu.edu/coca/).

All of the testing throughout the training was presented in a similar format. Students were

presented with a screen upon which two typed words were presented, a sound file played once

automatically and could be replayed one addition time but then had to make a selection.

Descriptive statistics for the results of the pretest on real words are reported in in Table 5-2.

Mean correct responses are out of 39 items for each group.

Table 5-2 Pretest on real words

| Real Words | Mandarin N=53 | Korean N=34 | Mandarin Control N=8 | Korean Control N=6 |
|---|---|---|---|---|
| Mean | 27.5 | 24.3 | 26.38 | 25.14 |
| Standard error | 0.57 | 0.73 | 1.49 | 1.59 |
| Median | 28 | 24 | 27 | 24 |
| Standard Dev. | 4.19 | 4.14 | 4.24 | 4.22 |
| Min. | 18 | 16 | 20 | 20 |
| Max. | 37 | 33 | 32 | 33 |

Since the pre-test scores on real words were not normally distributed, the Mann-Whitney

test statistic (Equivalent to Wilcoxon rank sum test) replaced a $t$-test to determine if there were

differences between the Mandarin and Korean students in accuracy rates on the real word pretest.

The Mandarin speaking group's scores were significantly higher than those of the Korean

speaking group ($t$ (81) = 3.50, $p < 0.01$). For Mandarin speakers, an ANOVA indicated no significant difference between scores for those assigned to different training paradigms, [$F(3,48)$ = 0.77, $p= 0.52$]. Within the Korean-speaking group, there were no difference between scores for participants assigned to different training paradigms [$F(3,27) = 2.34$, $p= 0.09$].

It was expected that the participants in this study would be more familiar with more frequent words, and would therefore be more accurate in identifying them. Figure 5-1 shows the relationship between accuracy and word frequency (plotted after converting to a log10scale for frequency scores) for those items on the real word pretest. Word frequency significantly predicted accuracy scores, $\beta= 22855.26$, $t(77) = 7.81$, $p < 0.01$, $R2 = .07$, $F(1,77) = 6.04$, $p=.014$.
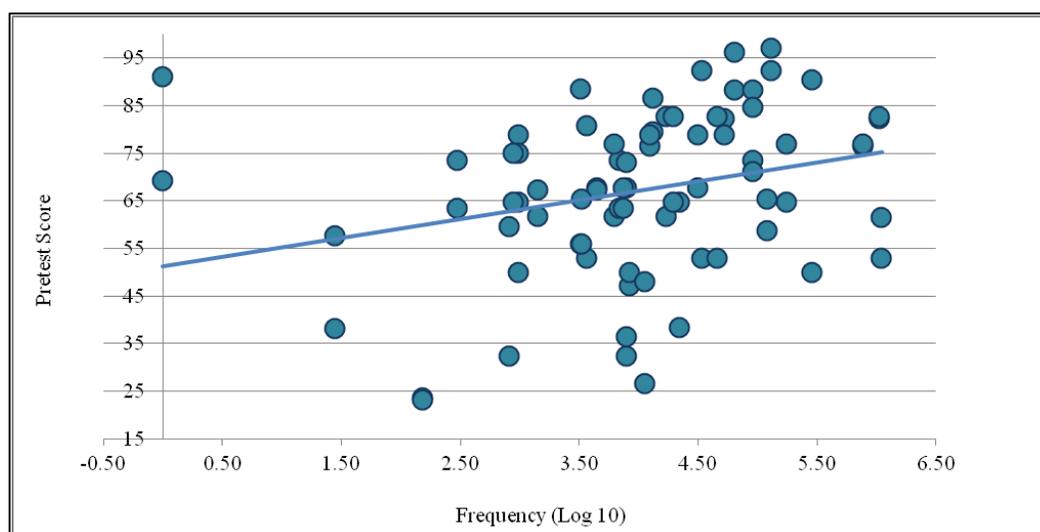


Fig. 5-1  Frequency by Accuracy
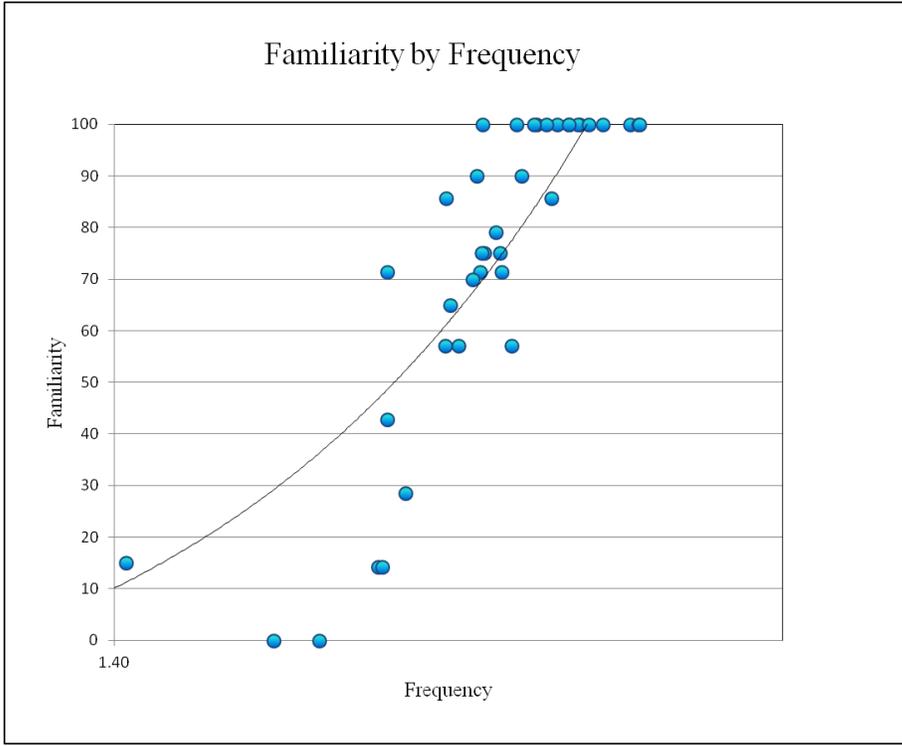
Fig. 5-2  Frequency by Familiarity

In addition to frequency ratings, participants completed a word familiarity survey, Figure 5-2 plots frequency (log 10) by familiarity scores (in percentages). Word frequency is a significant predictor of familiarity, [$F(1,38)=70.99$ $p<0.001$]. After completing the Real Word pretest, student-participants were trained in a classroom-based pronunciation exercise on the spelling conventions that were adopted for training; they then completed a nonce word pre-test. The results of the nonce-word pretest are presented in Section 5.3.

### 5.3 Pretest on Nonce Words

A nonce-word pretest was assigned as homework after in-class familiarization training on the spelling conventions. The pretest consisted of 30 items; six vowels were presented five times (or three minimal pairs were presented 10 times). Upon opening the screen, there was a one second delay and then the test item played automatically. There was a "play" button so that students could re-play any test item one additional time but then had to select one of the two options. Descriptive statistics for the results of the pretest on nonce words are reported in in Table 5-3. Mean correct responses are out of 30 items for each group.

Table 5-3 Average pre-test scores out of 30 nonce words

| Nonce Words | Mandarin N=53 | Korean N=34 | Mandarin Control (N=8) | Korean Control (N=7) |
|---|---|---|---|---|
| Mean | 21.25 | 19.97 | 21 | 20.83 |
| Standard Error | 0.35 | 0.569 | 1.16 | 1.13 |
| Median | 22 | 20 | 21 | 21 |
| Stndrd Dev. | 2.71 | 3.59 | 3.29 | 2.79 |
| Minimum | 15 | 11 | 16 | 18 |
| Maximum | 26 | 30 | 26 | 24 |

Korean speakers had more difficulty identifying vowels in nonce words ($M=19.82$, $SD=13.97$) than the Mandarin-speaking participants ($M=21.42$, $SD=7.09$) and the difference reached significance, $t(54)=2.15$, $p=0.04$. Within groups, a one way ANOVA indicated no significant difference between pretest scores on nonce words for Mandarin speakers assigned to the four different training paradigms, [$F(3,43) = 0.36$, $p= 0.78$]. Similarly, there was no significant differences between Korean speakers assigned to the four different paradigms,

$[F(3,30) = 0.07, p= 0.97]$. The overall distribution of scores on nonce words was normally distributed with the center of the distribution at 70% accuracy. Figure 5-2 displays the distribution of scores for nonce word identification. From this, it can be surmised that students learned the spelling conventions and the associated sounds such that the center of the distribution is at 70% accuracy.
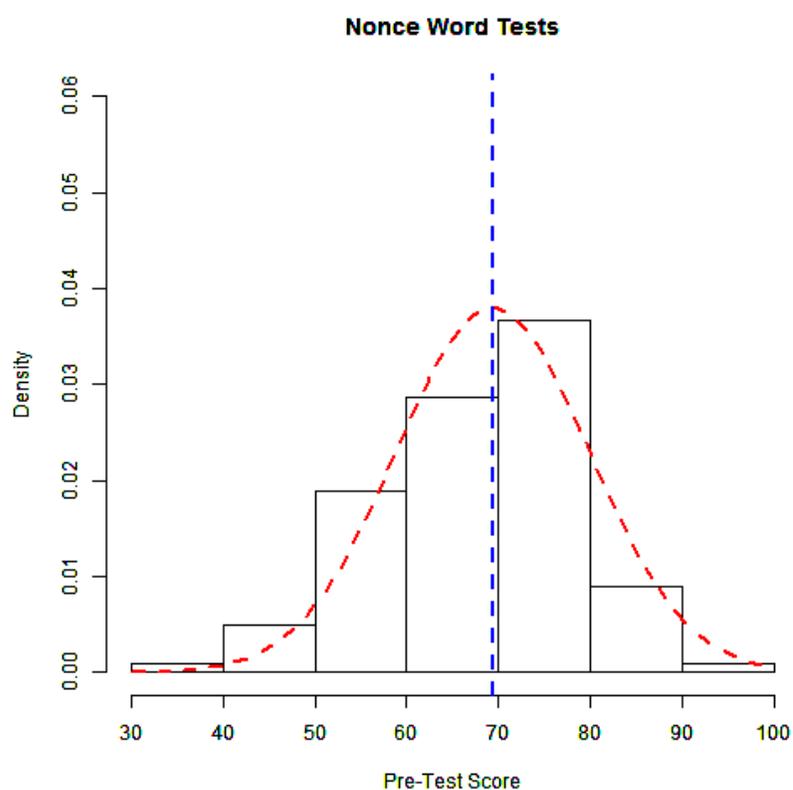


Figure 5-3 Distribution of pretest scores on nonce words follows a normal distribution

## 5.4 Nonce Word Pretest; Analysis by Vowel

Since the vowel inventory of the L1 may be implicated in accurate identification of the target vowels, performance of test takers on individual vowels is assessed. Percent accurate by

vowel and L1 is presented in table 5-4 and represented graphically in Figure 5-4. Two-Sample *t*-tests, assuming unequal variance, were conducted to compare vowel scores between L1-groups. Korean and Mandarin speakers differed significantly in accuracy rates identifying /i/, /ɪ/ and /u/. Identification of NAE /i:/ was significantly better for Mandarin speakers (*M*=4.30, *SD*=0.52) than Korean speakers (*M*=3.59, *SD*=1.28); $t(50)=3.27$, $p = 0.002$. Mandarin speakers were also significantly better at identifying /ɪ/ (*M*=4.09, *SD*=1.20) than their Korean classmates (*M*=3.32, *SD*= 1.07); $t(73)=3.31$, $p=0.001$. While both groups had more difficulty with /æ/, /ɛ/ and /ʊ/, scores were not significantly different between the two groups.

Table 5-4 Pretest average accuracy scores by vowel (N=5 per vowel) in percentages for Mandarin speakers (N=54) and Korean speakers (N=34) (* indicates significant differences between L1 groups.)

|       | æ     | ɛ     | i:     | ɪ      | u      | ʊ     |
|-------|-------|-------|--------|--------|--------|-------|
| M-Pre | 66.04 | 59.62 | *86.04 | *81.88 | *73.58 | 63.02 |
| K-Pre | 70.58 | 60.58 | 71.76  | 66.48  | 64.12  | 55.3  |
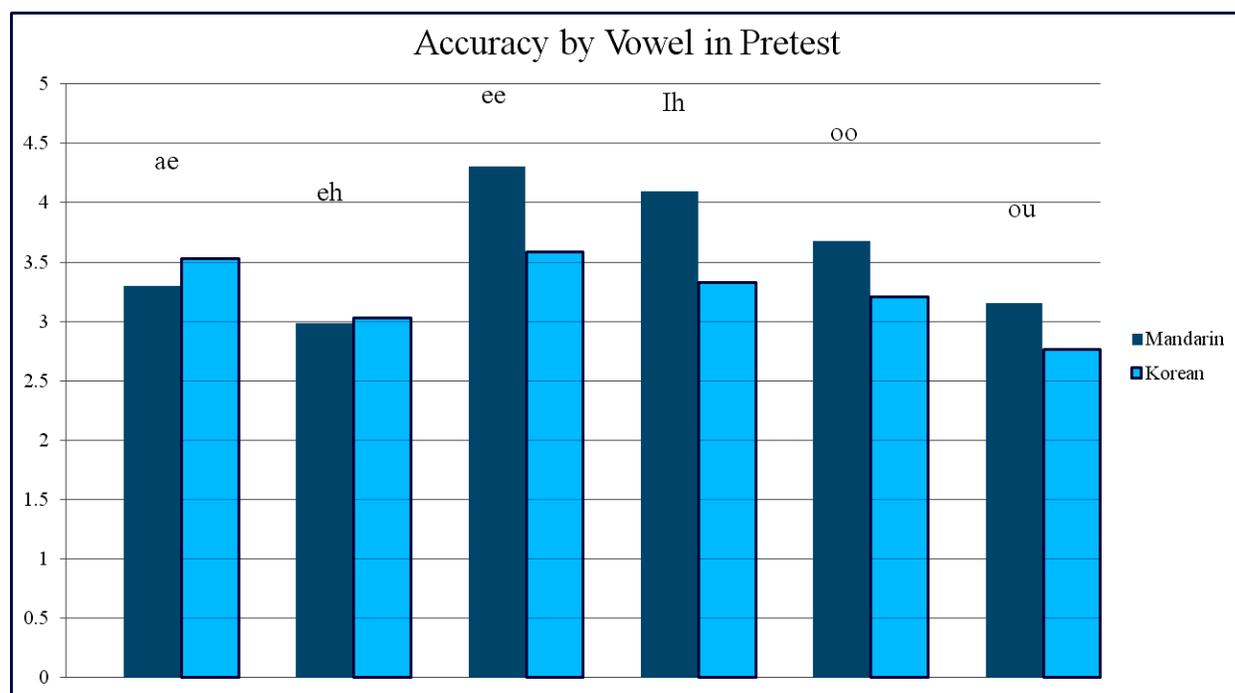
Fig. 5-4 Mandarin and Korean average raw score per vowel (N=5) in pretest

Given the vowel inventories of the L1, we also want to examine differences *within* each L1 group. A one-way, ANOVA was conducted to compare the accuracy scores on the identification of the six vowels for the Mandarin speakers; there was a significant variation between scores on vowels in the pretest [$F(5,312) = 10.93$, $p < 0.001$].

A post hoc Tukey multiple comparisons indicated that Mandarin scores on /i/ were significantly better than /æ/ ($p <0.001$); /ɛ/ ($p <0.001$); and /ʊ/ ($p <0.001$). Scores on /ɪ/ were significantly better than /æ/ ($p =0.01$); /ɛ/ ($p <0.001$); /ʊ/ ($p <0.001$) and /u/ was significantly better than /ɛ/ ($p =0.03$). No other differences between vowel identification were significant. For convenience, Mandarin pretest scores by vowel are repeated in Table 5-5 (values are averages out of five items for 54 participants):

Table 5-5 Mandarin participants' (N=54) accuracy scores by vowel (5 tokens for each vowel)

|  | iː | ɪ | u | æ | ʊ | ɛ |
|---|---|---|---|---|---|---|
| M-Pre | 4.302 | 4.094 | 3.679 | 3.302 | 3.151 | 2.981 |

Similarly, for the Korean L1 group, a one-way ANOVA was conducted to compare the accuracy scores on the identification of the six vowels for the Korean speakers; there was a significant difference between vowels on the pretest [$F(5,198) = 2.30$, $p = 0.046$].

A post hoc Tukey multiple comparisons indicated that Korean speakers' scores for /i/ was significantly better than /ʊ/ ($p$=05) identification. No other differences between vowels were significant. For convenience, accuracy scores by vowel for the Korean participants are repeated in Table 5-6.

Table 5-6 Korean participants' (N=34) average accuracy scores by vowel (5 tokens for each vowel)

|  | i | æ | ɪ | u | ɛ | ʊ |
|---|---|---|---|---|---|---|
| K-Pre | 3.588 | 3.529 | 3.324 | 3.206 | 3.029 | 2.765 |

## 5.5 Training

After completing the pretest on real words and nonce words, participants were assigned the task of completing a training session, followed by a test on a subset of the training materials, each day, for six days. The training behavior of participants is measured in terms of the total times they clicked on a "play" button to listen to a nonce word. Mandarin and Korean speakers

did not differ significantly in the amount of training; the Mandarin group played an average of

811 sound files (*SD*=621.96) and the Korean group averaged 831 (*SD*=81.98).

Since one of the research questions is the effect of variation on learning, and the amount

or type of variation in the training material may affect training behavior, it is worth exploring

differences between the training paradigms. Amount of training is measured by the number of

times a participant played the sound files. The averages by Paradigm and L1 are presented in

Table 5-7 and graphically in Figure 5-5.

Table 5-7 Mandarin and Korean training averages by paradigm

| Paradigm | *Mandarin* | *Korean* |
|----------|-----------|----------|
| A | 652.71 | 885 |
| B | 862.4 | 954.6 |
| C | 751.67 | 963.1 |
| D | 787 | 487.4 |

Figure 5-5 Amount of Training per Paradigm for Mandarin and Korean participants

Although the total amount of training was not significantly different between L1 groups, there is evidence from the bar graphs in Figure 5-4 that there were differences by paradigm. A one-way, between subjects ANOVA was conducted to compare the effect of paradigm membership on training totals in four conditions. For the Mandarin speakers, differences in training between paradigms were not significant; for the Korean speakers, however, there was a significant effect of paradigm membership on training totals for the four conditions [$F(3,200)$ = 11.86, $p < 0.001$]. A post hoc Tukey multiple comparison indicates that the average training for Korean speakers in Paradigm D was significantly *less* than those in Paradigm A ($p < 0.001$), B ($p < 0.001$) and C ($p < 0.001$). No other pair-wise comparisons indicated any significant differences.

Given differential accuracy on vowel identification in the pre-test, one would assume that learners would target those vowels that were the most difficult in the pretest. Average training totals by vowel for Mandarin and Korean speaking participants are presented in Table 5-8 and 5-9, respectively, and represented graphically in Figure 5-6.

Table 5- 8 Mandarin speakers' average total training by vowel across all training paradigms

| Mandarin | | | | | |
|---|---|---|---|---|---|
| *ae* | *eh* | *ee* | *ih* | *oo* | *ou* |
| 135.80 | 133.57 | 105.55 | 113.82 | 131.06 | 142.71 |

Table 5-9 Korean speakers' average total training by vowel across all training paradigms

| Korean | | | | | |
|---|---|---|---|---|---|
| *ae* | *eh* | *ee* | *ih* | *oo* | *ou* |
| 131.32 | 133 | 124.12 | 134.12 | 149.71 | 158.53 |
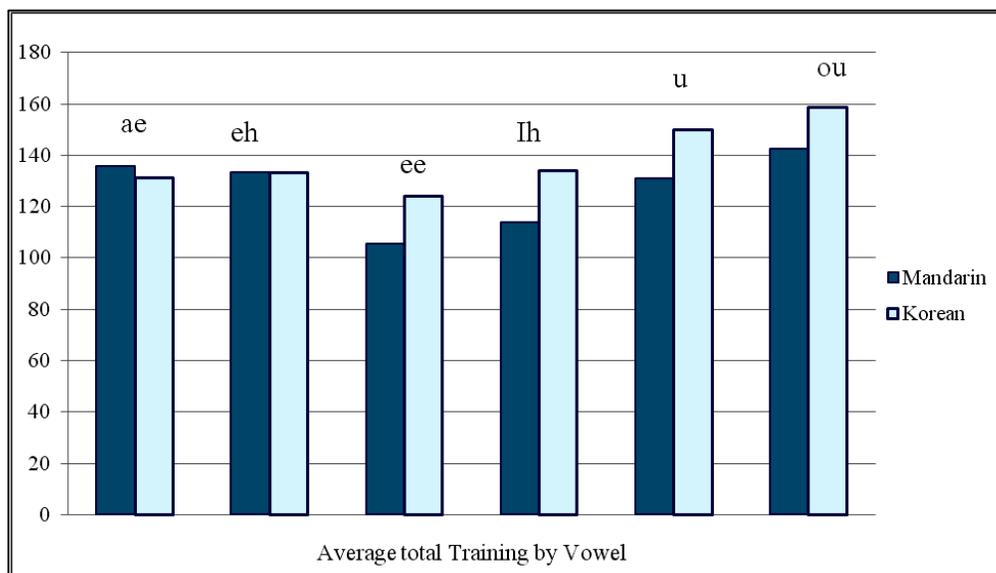


Figure 5-6 Korean and Mandarin speakers' average total training by vowel

Mandarin speakers trained the least on /i/ and /ɪ/ and both Korean and Mandarin speakers trained the most on /ʊ/. However, one-way ANOVA comparing training totals between vowels for Mandarin speakers indicated no significant differences, [$F(5,300)= 1.25. p=.28$]; this was similar for the Korean group, [$F(5,198)=.82, p=.54$]. Neither group trained on any particular vowel, or vowel pair, preferentially.

## 5.6  Real Word Post-tests

The following section reports post-test results for real words.  On the test of real word identification, Mandarin speakers, as a group, made significant gains between pre-test ($M=27.71$ $SD=17.23$) and post-test ($M=28.73, SD=17.85$) [$t(51) = -2.79, p=0.01$]. Korean speakers' as a group made no significant gains between pretest, $M=24.39, SD=18.05$, and post-test $M=25.03$, $SD=20.23, t(30) = 1.70, p=0.48$.

Considering differences by paradigm, Mandarin speakers in Paradigm A showed significantly greater accuracy in identifying real words in the post-test, ($M=30.21, SD=18.34$) than in pretest ($M=28.07, SD=18.84; t(13) = 1.77, p=0.01$). This was also the case for Mandarin participants in Paradigm B who made significant gains from pretest ($M=28.92, SD=15.41$ ) to post-test ($M=30.23, SD=14.03, t(12) = -2.34, p=0.04$). Korean participants in Paradigm A also showed significant improvement between pretest ($M=21.88, SD=7.27$) and post-test ($M=25.00$, $SD=10.29, t(7) = .55, p=0.02$) as did those Korean participants who trained in Paradigm B. Korean participants in Paradigm B averaged 23.29 ($SD=17.24$) correct responses out of 30 on the pretest and 25.86 ($SD=15.81$) on the post-test ($t(6) = -2.64, p=0.04$). Table 5-10 provides a summary of the pre- and post-test scores in percentages by L1 and training paradigm. Significant differences are in bold with an asterisk.

Table 5-10 Average percent correct of 39 tokens in real word (RW) post-test

|  | Paradigm | %PreRW | %PostRW | %Change |
|---|---|---|---|---|
| Mandarin | A | 71.98 | **77.47*** | 5.49 |
| Mandarin | B | 73.26 | **77.51*** | 4.25 |
| Mandarin | C | 69.63 | 68.44 | -1.18 |
| Mandarin | D | 68.16 | 70.73 | -2.56 |
|  | Paradigm | %PreRW | %PostRW | %Change |
| Korean | A | 56.09 | **64.10*** | 8.01 |
| Korean | B | 58.33 | **62.56*** | 4.23 |
| Korean | C | 64.36 | 63 | -1.36 |
| Korean | D | 68.27 | 66.67 | -1.6 |
|  |  |  |  |  |
| Mandarin | Control | 67.63 | 65.06 | -2.56 |
| Korean | Control | 64.47 | 65.57 | 1.1 |

The next section, Section 5.6, reports the results of post-test scores for nonce word identification for both Mandarin and Korean speakers and then by paradigm.

### 5.7 Nonce Word Post-tests

Both Mandarin and Korean participants made significant gains between pretest and post-test on nonce words. Mandarin speakers averaged 71.4% accurate (raw score $M$=21.42, $SD$=7.09) in the pretest and 82.1% (raw score $M$= 24.62, $SD$=6.24) on post-test ($t(52)$ = -8.92, $p$<0.001). Korean participants in pretest had a mean score of 66.6% (raw score $M$= 19.82, $SD$=13.97) and on post-test 77.1% (raw score $M$= 23.12, $SD$=9.62, $t(33)$= -5.39, $p$<0.001).

Considering scores by training paradigm, all groups made significant gains in identifying vowels in nonce words, save for Korean speakers in training paradigm D, the least variable condition ($t(7)$=-1.55, $p$=0.16) and the condition within which Korean participants trained the

least. Pre- and post-test scores and the percent gain for each L1 group, separated by training paradigm, is displayed in Table 5-11. Significant differences between pre- and post-test scores are in bold and indicated by an asterisk. The *t*-statistic and *p*-values by paradigm for Mandarin speakers are displayed in Table 5-12 and for Korean speakers in 5-13.

Table 5-11 Nonce word, post-test: Average percent accurate response out of 30 items

|  | Paradigm | %PreNonce | %PostNonce | %Change |
|---|---|---|---|---|
| Mandarin | A | 68.1 | **83.57*** | 15.47 |
| Mandarin | B | 74.05 | **84.76*** | 10.71 |
| Mandarin | C | 71.39 | **79.74*** | 8.35 |
| Mandarin | D | 70.3 | **79.72*** | 9.42 |
|  |  |  |  |  |
| Korean | A | 67.92 | **84.71*** | 16.79 |
| Korean | B | 65.83 | **75.42*** | 9.59 |
| Korean | C | 65.33 | **76.33*** | 11 |
| Korean | D | 65.42 | 71.25 | 5.83 |
|  |  |  |  |  |
| Mandarin | Control | 70 | 75.42 | 5.42 |
| Korean | Control | 69.44 | 78.89 | 3.81 |

Table 5-12. *t*-statistic and *p*-values by paradigm for nonce word pre- and post-score differences

| Mandarin | df | t stat | p |
|---|---|---|---|
| A | 13 | -5.20 | 0.001* |
| B | 13 | -7.87 | 0.001* |
| C | 12 | -3.13 | 0.01* |
| D | 11 | -4.02 | 0.002* |

Table 5-13. *t*-statistic and *p*-values by paradigm for nonce word pre- and post score differences

| Korean | df | t stat | p |
|--------|----|--------|-----|
| A | 7 | -4.16 | 0.004* |
| B | 7 | -2.76 | 0.03* |
| C | 9 | -2.63 | 0.03* |
| D | 7 | -1.55 | 0.2 |

For convenience, a summary table with all pretest, posttest, and gain scores for both real and nonce words broken down by L1 and control as well as by training paradigm are provided in Table 5-14. Asterisks indicate significant gains between pre- and post-test.

Table 5-14 Average accuracy scores and % change by Training Paradigm (* indicates statistically significant gains).

| | Paradigm | %PreRW | %PostRW | %Change | %PreNonce | PostNonce | %Change |
|---------|----------|--------|---------|---------|-----------|-----------|---------|
| Mandarin | A | 71.98 | **77.47*** | 5.49 | 68.1 | **83.57*** | 15.47 |
| Mandarin | B | 73.26 | **77.51*** | 4.25 | 74.05 | **84.76*** | 10.71 |
| Mandarin | C | 69.63 | 68.44 | -1.18 | 71.39 | **79.74*** | 8.35 |
| Mandarin | D | 68.16 | 70.73 | -2.56 | 70.3 | **79.72*** | 9.42 |
| | | | | | | | |
| Korean | A | 56.09 | **64.10*** | 8.01 | 67.92 | **84.71*** | 16.79 |
| Korean | B | 58.33 | **62.56*** | 4.23 | 65.83 | **75.42*** | 9.59 |
| Korean | C | 64.36 | 63 | -1.36 | 65.33 | **76.33*** | 11 |
| Korean | D | 68.27 | 66.67 | -1.6 | 65.42 | 71.25 | 5.83 |
| | | | | | | | |
| Mandarin | Control | 67.63 | 65.06 | -2.56 | 70 | 75.42 | 5.42 |
| Korean | Control | 64.47 | 65.57 | 1.1 | 69.44 | 78.89 | 3.81 |

The control group made some gains in nonce word scores – presumably from their participation in the classroom familiarization exercise – but these were not significant. Section 5.7 reports on gains, or in some cases losses, on individual vowels for Mandarin and Korean participants and by training paradigm.

## 5.8 Improvement by Vowel

Differences between pretest and post-test scores for individual vowels were investigated using paired t-tests for means. Table 5-15 summarizes scores for the Mandarin-speaking students by vowel in pre-and post-test with gains. Bold scores with asterisks mark significant gains.

Table 5-15 Mandarin % average gain (pre-test – post-test) by vowel

|          | æ      | ɛ       | i     | ɪ      | u      | ʊ       |
|----------|--------|---------|-------|--------|--------|---------|
| M-Pre    | 66.04  | 59.62   | 86.04 | 81.88  | 73.58  | 63.02   |
| M-Post   | 74.72  | 70.19   | 90.19 | 88.68  | 81.89  | 86.42   |
| % change | 8.68*  | 10.57*  | 4.15  | 6.80*  | 8.31*  | 23.40*  |

Mandarin speakers improved on their perception of every vowel except /i/ (*M*=4.30, *SD* =.52 in pretest, *M*=4.51, *SD*=.79 in post-test, *t*(52)=-1.35, *p*=.18). However, /i/ was also the most accurately identified vowel in the pretest and reached 90.19% accuracy in the post-test. The vowels /ɪ/ (*M*=4.09, *SD*=1.20) and /u/ (*M*=3.68, *SD*=1.49) which were relatively advantaged in the pretest, made significant improvement in the post-test (*M*=4.43, *SD*=.67, *t*(52)=-1.97, *p*=.05 and *M*=4.09, *SD*=1.05, *t*(52)=-2.45, *p*=0.02, respectively). Their greatest improvement was on /ʊ/ (*M*=3.15, *SD*=2.17 in pretest and *M*=4.32, *SD*=.68, *t*(52)=-5.72, *p*<0.001) which reached 86.42% accuracy in the post-test. The Mandarin group had the most difficulty with /ɛ/ in pretest (*M*=2.98,

*SD*=1.37) and although improvement was significant (*M*=3.51, *SD*=1.10, *t*(52)=-2.61, *p*=0.01),

Mandarin speakers only reached 70% accuracy in post-test identification of /ɛ/. This was

followed closely by /ɛ/'s contrasting partner /ae/ (in pretest, *M*=3.30, *SD*=1.48) which also

showed significant improvement (*M*=3.74, *SD*=1.24, *t*(52)= -2.02, *p*=0.05) but only reached 75%

accurate identification in the post-test.

Korean speakers made gains in the accurate identification of all vowels except for /ae/

(*M*=3.53, *SD*=1.48, *t*(33)=0, *p*=1.0) and /i/ (*M*=3.94, *SD*=1.09, *t*(33)=-1.58, *p*=0.12), the latter of

which was the most accurately identified vowel in pretest (*M*=3.59, *SD*=1.28). The contrasting

counterparts, /ɛ/ (*M*=3.03, *SD*=.64) and /ɪ/ (*M*=3.32, *SD*=1.07) made significant gains in post-test

(*M*=3.71, *SD*=.88, *t*(33)=-3.70, *p*< 0.001 and *M*=4.09, *SD*=.63, *t*(33)= -4.04, *p*=.001,

respectively). As with their Mandarin counterparts, the greatest improvement was on /ʊ/

(*M*=3.94, *SD*=.97, *t*(33)= -4.13, *p*<0.001). Although Korean participants made significant gains

in four of the six vowels, their scores only exceed 80% accuracy in instance, that of /ɪ/

identification. Pre-and post-test scores by vowel for the Korean participants are presented by

vowel in Table 5-16; significant gains are marked in bold with an asterisk.

Table 5-16 Korean % average gain (pre-test – post-test) by vowel

|          | æ     | ɛ      | i     | ɪ      | u      | ʊ      |
|----------|-------|--------|-------|--------|--------|--------|
| K-Pre    | 70.58 | 60.58  | 71.76 | 66.48  | 64.12  | 55.3   |
| K-Post   | 70.58 | 74.12  | 78.82 | 81.76  | 76.48  | 78.82  |
| % change | 0     | **13.54*** | 7.06  | **15.28*** | **12.36*** | **23.52*** |

The bar graph in Fig. 5-7 represents pre- and post-test scores by group and by vowel. Patterns of improvement are clear across vowel categories – save for the Korean learners for /æ/ for which there was no improvement.
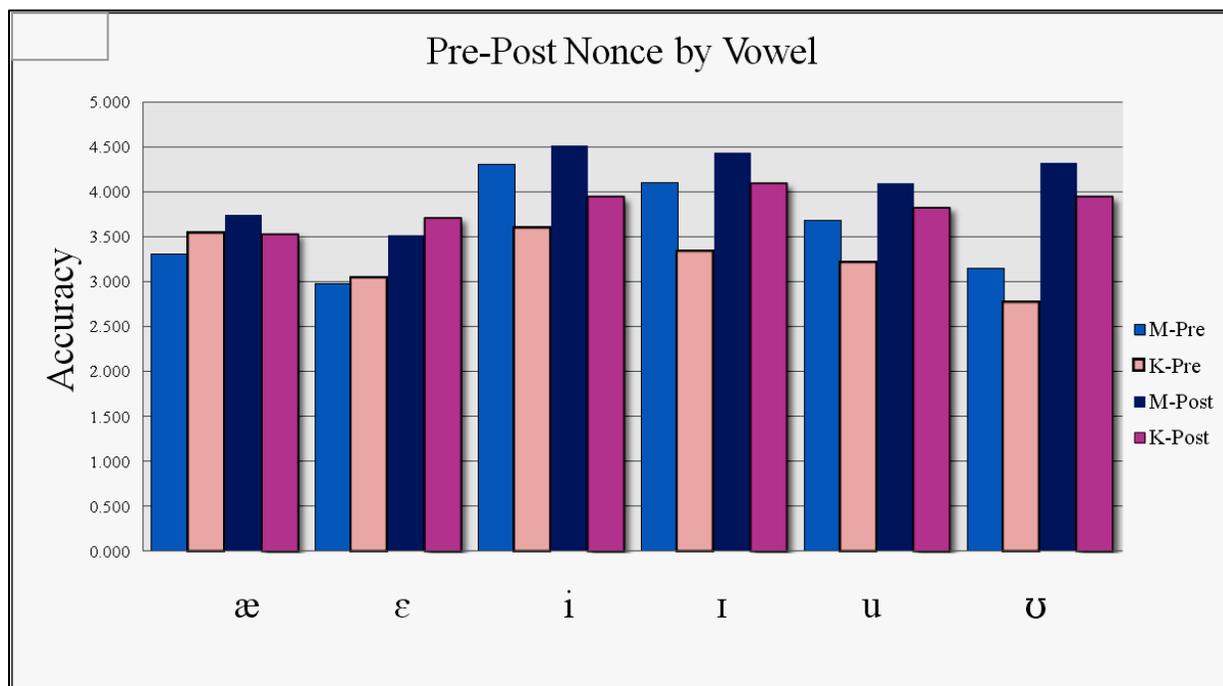


Fig. 5-7 Mandarin and Korean gains by vowel (raw accuracy scores are out of 5)

Although improvement in vowel discrimination by L1 is encouraging, of equal interest is how gains are distributed between training paradigms. Section 5.3.2 reports on differences between pre- and post-test scores by vowel for each training paradigm.

**5.9 Gains by Vowel by Paradigm**

Both groups, regardless of training paradigm, made gains in their accuracy at identifying certain vowels. However, of central interest is how individuals improve on different vowels as a

function of the training paradigm to which they were assigned. In Table 5-17, Mandarin

speakers' percent improvement by vowel by paradigm is presented. Significant differences

between pretest and post-test are indicated in bold with an asterisk.

Table 5-17 Mandarin: Gain scores (Poste-test-Pre-test)  by vowel by paradigm

| Vowel | Paradigm | | | |
|---|---|---|---|---|
| | A | B | C | D |
| æ | **18.57*** | 1.43 | 6.15 | 8.33 |
| ɛ | -2.86 | **25.71*** | 12.31 | 6.67 |
| i | **14.29*** | 2.86 | 4.62 | -6.67 |
| ɪ | 11.43 | 2.86 | 1.54 | **11.67*** |
| u | **17.14*** | 12.86 | 7.69 | -6.67 |
| ʊ | **20*** | **17.14*** | **27.69*** | **30.00*** |

Mandarin speakers improved significantly on their identification of /ʊ/ regardless of

paradigm . Those Mandarin-speaking participants assigned to Paradigm A made significant gains

on /æ/, /i/, /u/, and /ʊ/. Participants in Paradigm B made significant gains on /ɛ/ in addition to /ʊ/

with participants in Paradigm C only gaining on /ʊ/. Participants in Paradigm D made significant

gains in identifying /ɪ/ and /ʊ/. For ease of readability, related statistics are presented in Table 5-

18.

Table 5-18. Mandarin: summary statistics for differences between pre- and post-test score for vowels by paradigm

| Paradigm | Vowel | Pre *M* | Pre *SD* | Post *M* | Post *SD* | *t*-stat | *p* value |
|----------|-------|---------|----------|----------|-----------|----------|-----------|
| A | ae | 3.14 | 0.9 | 4.07 | 0.53 | *t*(13)= -2.88 | *p*=0.013 |
| A | ee | 4.07 | 0.53 | 4.79 | 0.18 | *t*(13)= -2.92 | *p*=0.012 |
| A | oo | 3.43 | 1.19 | 4.29 | 0.68 | *t*(13)= -2.92 | *p*=0.012 |
| A | ou | 3.29 | 2.07 | 4.29 | 0.37 | *t*(13)= -2.65 | *p*=0.020 |
| B | eh | 3 | 1.54 | 4.29 | 0.37 | *t*(13)= -3.63 | *p*=0.003 |
| B | ou | 3.79 | 0.64 | 4.64 | 0.55 | *t*(13)= -3.71 | *p*=0.003 |
| C | ou | 3.08 | 1.91 | 4.46 | 0.27 | *t*(12)= -3.77 | *p*=0.003 |
| D | Ih | 4.25 | 0.57 | 4.83 | 0.15 | *t*(11)= -2.55 | *p*=0.027 |
| D | ou | 2.33 | 3.7 | 3.83 | 1.42 | *t*(11)= -2.32 | *p*=0.040 |

Korean-speaking participants gains by vowel by paradigm are summarized in Table 5-19 with significant gains indicated with an asterisk. The related statistics are presented in Table 5-20. Korean participants in Paradigm A made significant gains in identifying /ɛ/. No other vowels showed a significant difference between pre- and post-test in Paradigm A. The vowel /æ/ made no significant gains regardless of paradigm. Significant gains were made for /ɛ/, /i/, /ɪ/, /u/, and /ʊ/ but only in Paradigm C.

Table 5-19 Korean: Gain scores by vowel by paradigm

| Vowel | Paradigm | | | |
|-------|------|------|--------|------|
| | A | B | C | D |
| æ | 10 | 10 | 0 | -20 |
| ɛ | **22.50*** | 7.5 | **16.00*** | 7.5 |
| i | 10 | 5 | **14.00*** | -2.5 |
| ɪ | 17.5 | 17.5 | **14.00*** | 12.5 |
| u | 7.5 | 0 | **22.00*** | 17.5 |
| ʊ | 25 | 17.5 | **28.00*** | 22.5 |

Table 5-20.Korean: summary statistics for differences between pre- and post-test score for vowels by paradigm

|   | Vowel | Pre *M* | Pre *SD* | Post *M* | Post *SD* | *t* stat | *p* value |
|---|---|---|---|---|---|---|---|
| A | ɛ | 2.63 | 0.84 | 3.75 | 0.50 | *t* (7)= -2.83 | *p*=0.026 |
| C | ɛ | 3.20 | 0.18 | 4.00 | 0.67 | *t* (9)= -2.75 | *p*=0.022 |
| C | i | 3.50 | 2.06 | 4.20 | 0.40 | *t* (9)= -2.33 | *p*=0.045 |
| C | ɪ | 3.40 | 0.71 | 4.10 | 0.32 | *t* (9)= -2.33 | *p*=0.045 |
| C | u | 2.90 | 1.43 | 4.00 | 0.67 | *t* (7)= -2.54 | *p*=0.032 |
| C | ʊ | 2.40 | 1.38 | 3.80 | 1.07 | *t* (7)= -1.47 | *p*=0.013 |

Gains, or in some cases loses, are presented graphically in Fig. 5-8 through 5-11. Figures 5-8 and 5-9 represent the training paradigm along the x-axis with the vowels as reponse variables plotted on the y-axis. Figure 5-10 and 5-11 present the same information but with the vowels plotted on the x-axis and the pardigm on the y-axis.
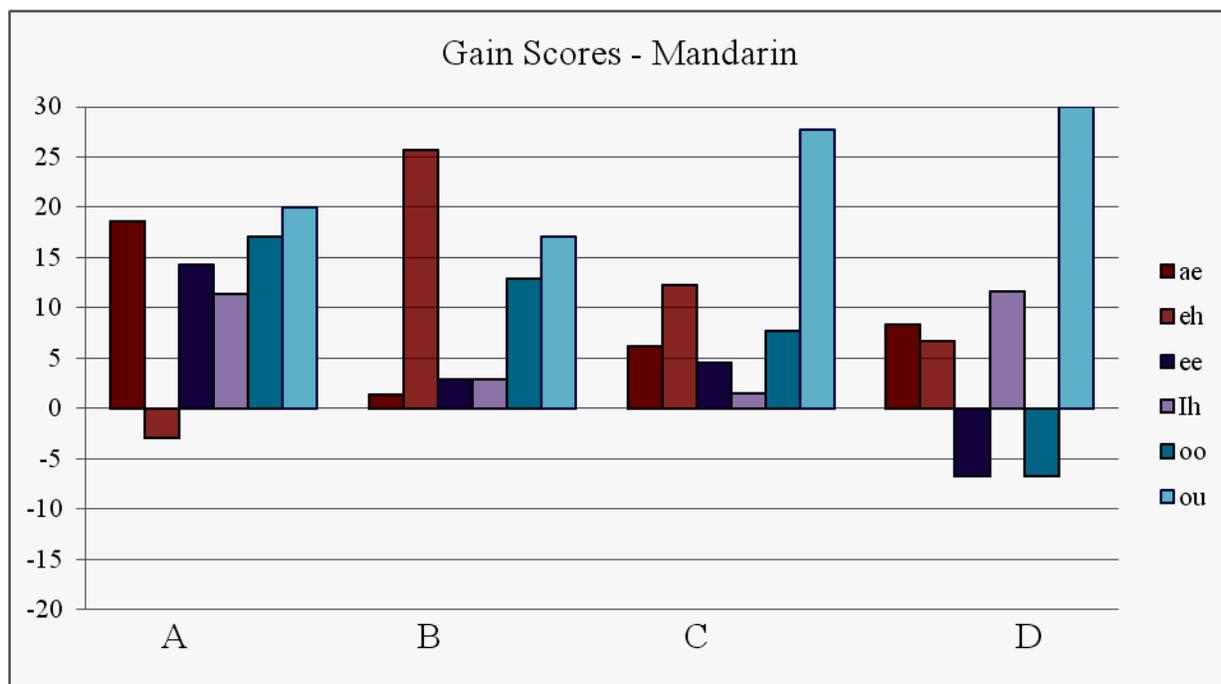
Figure 5-8 Mandarin: Average percent gain (y-axis) by paradigm (x-axis), by vowel.
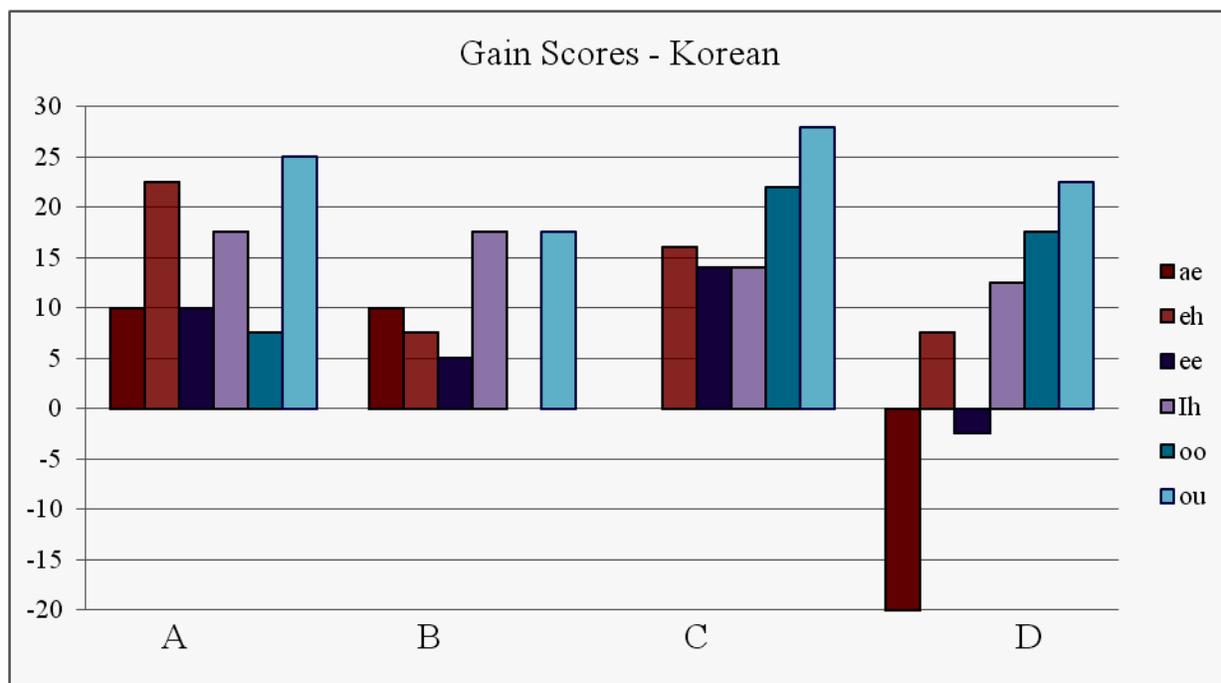


Figure 5-9 Korean: Average percent gain (y-axis) by paradigm (x-axis), by vowel.
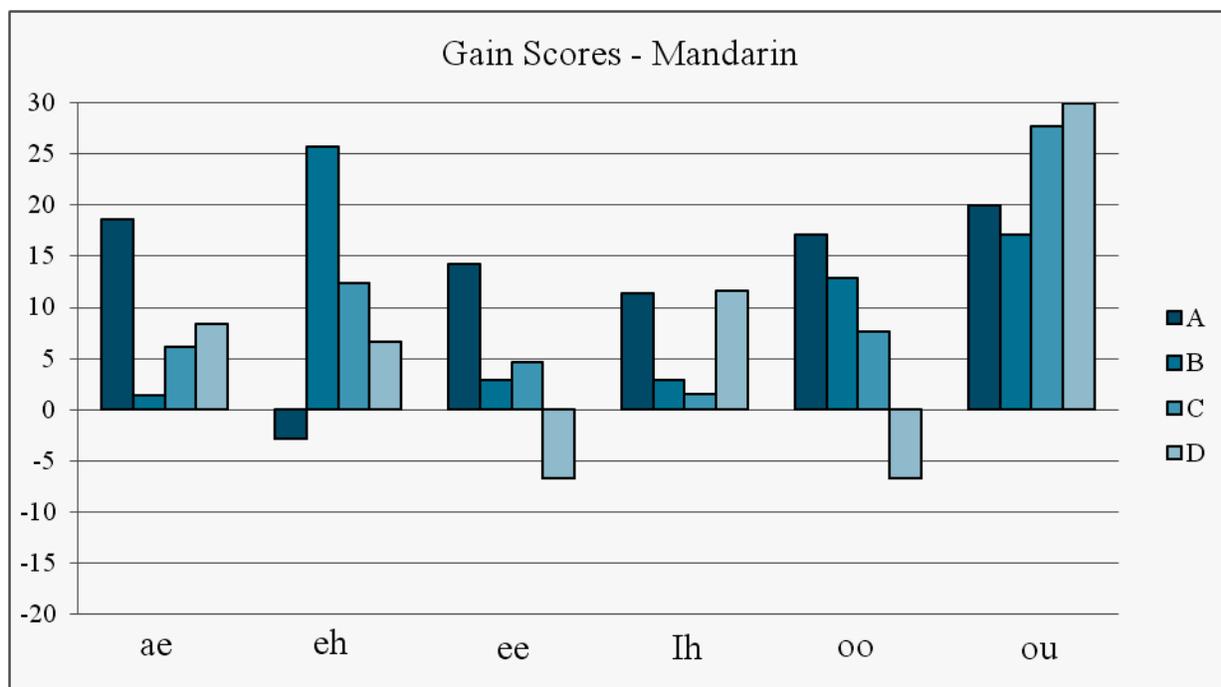
Figure 5-10 Mandarin: Average percent gain (y-axis) by vowel (X-axis) by paradigm
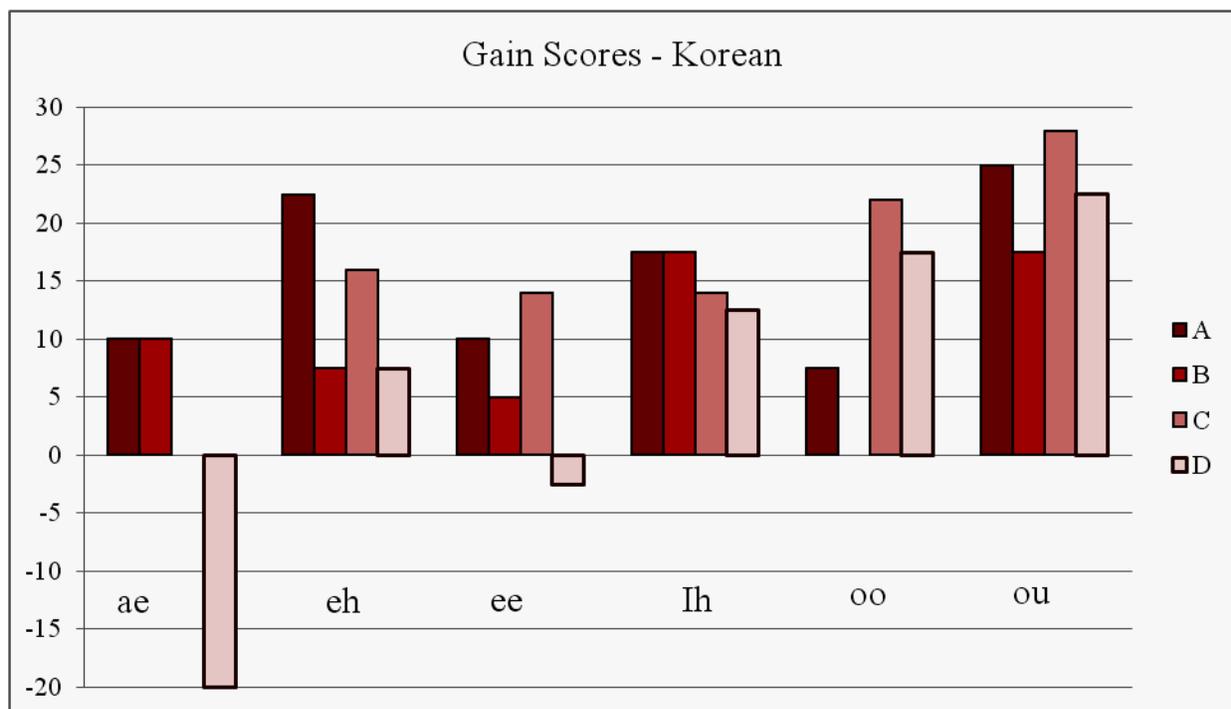
Figure 5-11 Korean: Average percent gain (y-axis) by vowel (X-axis) by paradigm

Given that there were no signficant differences between training on any particular vowel, nor were there significant differences between training by paradigm (save for Korean participants in Paradigm D), then differences in improvement suggests that differences in the training paradigms has an effect on vowel learning. For the Mandarin speakers, improvement in /æ/, /ɛ/, /i/ /u/ improved under conditions with multiple speakers. Only /ɪ/ made significant improvement in Paradigm D, the least variable condition, except for /ʊ/ which improved under all training conditions. For the Korean speakers, who were significantly less accurate in pretest for both real and nonce words, Paradigm C, one speaker but with increasing speech rates, was clearly the best training condition for the majority of vowels. In the pretest, Korean speakers had difficulty with /ae/ and this vowel did not significantly improve regardless of training paradigm. Only /ɛ/ improved under two Paradigms, both A and C.

Another area of variation, that of coda condition, is explored in the following section.

## 5.10 Coda condition

Mandarin has a relatively restricted syllable inventory CGVC and CGVG where coda consonants are limited to /n/ and /ŋ/. Korean's syllable inventory is (C)(G)V(C) and codas are limited to /p/, /t/, /k/, /m/, /n/, /ŋ/, and /l/, (Ha et al., 2009). Recall that the nonce items were constructed with onsets and codas /p/, /t/, /k/, /b/, /d/, /g/, /m/, /n/, /s/, /h/ where /h/ is restricted to onsets and /s/ to codas.

### 5.10.1 Errors and codas in the nonce words pretest

 In the pre-test on nonce words, Mandarin speakers identified just a little over half, 16 of 30 (53%) of the vowels with 70% accuracy and above. At post-test, 23 out of 30 tokens (73%) were identified at 70% accuracy and above. For the Mandarin speakers, of the eight nonce words in the post-test that were problematic, that is, identified accurately 69% of the time or less, five codas were nasals, one was a voiced velar stop, /g/, and one /s/ and one /b/. Nasal codas are allowed in Mandarin although /g/, /s/ and /b/ are not.

Korean learners were similarly positioned; 14 of 30 vowels were identified with 70% accuracy and above in the pre-test and at post-test 21 tokens out of 30 were identified at 70% accuracy and above. For the Korean speakers, of the nine tokens in the post-test that were still difficult and identified at 69% or less, three codas were nasals, three were voiced velar stops, /g/, two were /s/ and one was a /b/. Korean allows nasal codas; /g/ and /b/ are not allowed although their voiceless counterparts are. Korean does not allow /s/ in coda position.

Given the differences in licit codas between Korean and Mandarin, the similar difficulty with the same codas is surprising. The errors associated with nasals are particularly perplexing, given that both languages allow nasals in codas. Although in Mandarin, /m/ is restricted to onsets, the effect on the preceding vowel should not be different. In Mandarin, however, nasal codas vary by context; /ŋ/ follows the low back vowel /ɑ/ and /n/ follows the low front vowel /a/. In general, nasal place of articulation co-varies with vowel backness (Li, 2008). So the fact that Mandarin allows nasals but that they interact with the backness of the preceding vowel may add another layer of interference in vowel identification. In order to explore the interaction of vowel identification with the coda condition, errors from the daily post-training tests were examined.

**5.10.2 Errors and codas: daily training and post-tests - Mandarin**

During each day of training, participants could listen to as many sound files as they wanted. When they chose to finish, there was a test of 25 items selected from that days training material. An ANOVA was run and indicated a significant effect for coda [$F(40,54)=2.10$, $p=0.005$]. In Table 5-22 is a sampling of 849 errors made by Mandarin speakers from the six tests. The vowels are listed in the left column with total errors for that vowel adjacent. Codas are arranged along the following columns.

From Table 5-22, there is evidence that the coda condition differentially affects the accuracy of vowel identification. For example, for the Mandarin speakers, the nasal coda /n/ has the greatest detrimental effect on lax vowels in general. The nasal coda /m/ however, has a greater impact just on /æ/~/ɛ/. It's interesting to note that /i/ has the least errors throughout the testing, as one might expect given the accurate identification rates in the pretest. Pretest scores by vowel for the Mandarin group is repeated in Table 5-21 for convenience.

Table 5-21- Mandarin average pre-test scores, in percentages, by vowel (N=5)

|  | æ | ɛ | i | ɪ | u | ʊ |
|---|---|---|---|---|---|---|
| M-Pre | 66.04 | 59.62 | 86.04 | 81.88 | 73.58 | 63.02 |

Note that errors with /i/ identification are disproportionately associated with /p/ and /s/ in the coda. In the pretest, /ɪ/ was also significantly better identified than other vowels but does not enjoy the same advantage during the course of training; error rates for /ɪ/, 16.25%, have the same percentages as the error rates for /æ/, 16.61%, and /ʊ/ at 16.73%. Another one of the better identified vowels in the pretest, /u/, accounts for 12.84% of the errors throughout the training-testing. However, /u/ seems to be more adversely affected by voiceless codas than voiced coda. Errors by feature, voiceless stops, voiced stops, /s/ (the only voiceless sibilant) and nasals are presented in Table 5-23. Of the errors identifying /u/, 40.37% are in the context of /p,t,k/; compare this to /b,d,g/ which account for only 14.68% of errors. An /s/ in coda position contributes nearly 25% to the total errors identifying /u/. Both /u/ and /i/, both high, tense vowels, are disproportionately misidentified when /s/ is the coda.

Table 5-22- Sampling of 849 errors from daily testing for Mandarin speakers.

|  | Total% | p | t | k | b | d | g | s | m | n |
|---|---|---|---|---|---|---|---|---|---|---|
| æ | 16.61 | 18.44 | 9.22 | 0.00 | 7.80 | 10.64 | 26.95 | 2.84 | 23.40 | 0.71 |
| ɛ | 27.92 | 5.06 | 4.64 | 5.49 | 20.25 | 18.57 | 3.38 | 5.91 | 16.03 | 20.68 |
| i | 9.66 | 39.02 | 8.54 | 2.44 | 6.10 | 1.22 | 0.00 | 26.83 | 8.54 | 7.32 |
| ɪ | 16.25 | 7.97 | 10.14 | 13.77 | 10.87 | 12.32 | 18.84 | 2.90 | 3.62 | 18.12 |
| u | 12.84 | 17.43 | 11.93 | 11.01 | 6.42 | 3.67 | 4.59 | 24.77 | 8.26 | 11.93 |
| ʊ | 16.73 | 5.63 | 9.15 | 11.97 | 8.45 | 7.75 | 27.46 | 3.52 | 8.45 | 17.61 |

Table 5-23. Mandarin 849 errors by codas collapsed into classes.

| Mandarin | voiceless | voiced | /s/ | nasals |
|----------|-----------|--------|-------|--------|
| æ | 27.66 | 45.39 | 2.84 | 24.11 |
| ɛ | 15.19 | 42.19 | 5.91 | 36.71 |
| i | 50 | 7.32 | 26.83 | 15.85 |
| ɪ | 31.88 | 42.03 | 2.9 | 21.74 |
| u | 40.37 | 14.68 | 24.77 | 20.18 |
| ʊ | 26.76 | 43.66 | 3.52 | 26.06 |

## 5.10.3 Errors and codas: daily training and post-tests - Korean

Similarly, Korean speakers' vowel errors were also affected by coda condition. Pretest scores for the Korean speaking students are repeated in Table 5-24 for convenience.

Table 5-24-Korean average pre-test scores, in percentages, by vowel (N=5)

|  | æ | ɛ | i | ɪ | u | ʊ |
|-------|-------|-------|-------|-------|-------|-------|
| K-Pre | 70.58 | 60.58 | 71.76 | 66.48 | 64.12 | 55.3 |

In the pretest, Korean speakers' errors were relatively evenly distributed in terms of vowel identification with the only significant difference between /i/ and /ʊ/. In Table 5-25, a sampling of 1,436 errors on vowel identification taken from the post-training tests during the six days of training for the Korean speaking group are presented. As with the pretest scores, error rates were relatively evenly distributed between vowels. Similar to their Mandarin counterparts, an /m/ in the coda impacts the identification of /ae/ more than other vowels.

Table 5-25: Sampling of 1,436 errors from daily testing for Korean speakers.

| | Total% | p | t | k | b | d | g | s | m | n |
|---|---|---|---|---|---|---|---|---|---|---|
| æ | 15.88 | 11.4 | 2.63 | 0 | 11.84 | 21.49 | 18.86 | 6.14 | 26.75 | 0.88 |
| ɛ | 14.76 | 12.26 | 15.57 | 4.72 | 15.57 | 10.38 | 1.42 | 0.94 | 16.51 | 22.64 |
| i | 16.99 | 25.41 | 16.39 | 11.48 | 16.39 | 7.38 | 0 | 11.89 | 5.74 | 5.33 |
| ɪ | 17.27 | 9.68 | 10.89 | 18.15 | 16.94 | 9.27 | 10.48 | 10.08 | 6.45 | 8.06 |
| u | 17.34 | 12.85 | 7.23 | 8.84 | 11.24 | 17.67 | 6.43 | 20.88 | 6.43 | 8.43 |
| ʊ | 17.76 | 12.94 | 20.39 | 20.39 | 7.45 | 6.67 | 8.63 | 8.24 | 5.49 | 9.8 |

Table 5-26 collapses coda errors into consonant features, voiceless stops, voiced stops, /s/, a voiceless sibilant, and nasals. Although Korean allows voiceless consonants in codas, they account for just over 53% of the errors for /i/ and nearly 54% of the errors on /ʊ/. Voiced codas on the other hand, contribute 52.19% of errors on /æ/ versus only 14.04% for voiceless codas. While coda /m/ differentially impacted the identification of /ae/, collapsing /n/ and /m/ into one category reveals that of all the errors associated with a nasal coda, 62.78% of those errors are in identifying /æ/ and /ɛ/.

Table 5-26. Korean 1,436 errors by codas collapsed into classes.

| Korean | voiceless | voiced | s | nasals |
|---|---|---|---|---|
| æ | 14.04 | 52.19 | 6.14 | 27.63 |
| ɛ | 32.55 | 25.36 | 0.94 | 35.15 |
| iː | 53.28 | 23.77 | 11.89 | 11.07 |
| ɪ | 38.71 | 36.69 | 10.08 | 14.52 |
| u | 28.92 | 35.34 | 20.88 | 14.86 |
| ʊ | 53.72 | 22.75 | 8.24 | 15.29 |

In summary, Korean and Mandarin speakers who trained in Paradigm A and B made significantly better scores on real-word post-test than on the pretest. Mandarin speakers, regardless of training paradigm, made significant gains identifying vowels in nonce words in the post-test. Korean speakers made significant gains in Paradigm A. B. and C but not in D. By individual vowel, Mandarin speakers who trained in Paradigm A, improved on four, /æ, i, u, ʊ/, of six vowels. Of those four vowels, /i/ and /u/ had relatively high accuracy scores in pretest while /æ/ and /ʊ/ were two of the more difficult vowels. In Paradigm B, Mandarin speakers improved on their ability to identify /ɛ/ and /ʊ/ which were among the more difficult vowels to identify in the pre-test. For those who trained in Paradigm C, only /ʊ/ improved and in Paradigm D both /ʊ/ and /ɪ/ improved. Although /ʊ/ was one of the more difficult vowels in pre-test, /ɪ/ was identified at rates significantly higher than all other vowels. However, in the examination of errors across all training, /ɪ/ appeared to be one of the more difficult vowels. Korean speakers did not perform as well as their Mandarin counterparts on pretest for either real or nonce words, although they made significant gains on real word identification if they trained in Paradigm A or B. On nonce words, if Korean participant trained in Paradigm A, B or C, he/she made significant gains on the post-test. Only those in Paradigm D failed to make any significant improvement. By individual vowel, /ae/ was difficult for the Korean speakers in the pretest and made no significant gains regardless of paradigm. In Paradigm A, again the highest variability training paradigm, Korean speakers had significant differences between pre- and post-test on the identification of /ɛ/ also one of the more difficult vowels in pretest. All other significant differences between pre- and post-test for identifying vowels /ɛ, i, ɪ, u/ and /ʊ/ were for those Korean participants who trained in Paradigm C.

## 5.11 Improvement by Pre-test Score

Another area of interest is related to differences between participants in terms of starting levels of proficiency, training and improvement. Table 5-27 summarizes average percent improvement and average total training by pretest scores divided into three groups, those with initially low, medium and high scores. Given the range of the data, average target scores of 60% were considered "low", 70% "medium" and 80% "high". Participants with the highest and lowest pretest scores trained the most compared to those with scores in the median range. Examining improvement by paradigms, presented in Table 5-28, those with low pretest scores improved the most in Paradigm A and C; those in the mid-range improved the most in Paradigm A and D and those with the highest scores improved the most in Paradigm A and B.

Table 5-27. Training and improvement by pretest scores (Percent average correct out of 30).

|  | # of learners | Avg. Pretest | Avg. Improveme | Avg total Training |
|---|---|---|---|---|
| Low (37% - 60%) | 21 (24%) | 55% | 18% | 1,389 |
| Medium (63%-73%) | 39 (45%) | 69% | 11% | 811 |
| High (77% - 100%) | 27 (31%) | 81% | 4.60% | 1,032 |

Table 5-28. Gain (post-test – pre-test) by paradigm.

| Avg. Improvement | A | B | C | D |
|---|---|---|---|---|
| Low (37% - 60%) | 27.50% | 16.70% | 18.30% | 5.60% |
| Medium (63%-73%) | 14.80% | 8.70% | 8.30% | 9.60% |
| High (77% - 100%) | 6.70% | 7.90% | -1.43% | 5% |

While promising that most individuals improved between pre-and post-test, the relationship between L1, beginning proficiency, and what kind of variation impacts training, is not straightforward.

In the next chapter, Chapter 6, the results are discussed in terms of the research questions as well as the hypotheses about possible outcomes.

## Chapter 6: Discussion

### 6.1 Pretests: Real and Nonce Words Tests

In the Real Word pretest participants Mandarin speakers performed better than their Korean classmates. Although participants achieved a mean score of 70% accuracy on the pretest on nonce words, Mandarin speakers also out-performed their Korean counterparts. From these initial results, it must be concluded that the participants are at different proficiency levels – at least in terms of their respective abilities to identify vowels.

### 6.2 Pretest: Performance on Vowels

Mandarin speakers were significantly better identifying vowels /i, ɪ, u/ than the Korean speaking students. Both groups had more difficulty with /æ, ɛ & ʊ/ but scores were not significantly different between the two groups.

Within groups, Mandarin scores for /i/ and /ɪ/ identification were significantly better than for all other vowels. Accurate identification of /u/ was significantly better than /ʊ/ and /ɛ/. Korean speakers were significantly better at identifying /i/ than /ɛ/, and /i/ than /ʊ/ but there were no other significant differences between vowel-pair identification scores. Pretest scores by vowel for Mandarin and Korean participants are repeated in Table 6-1.

Table 6-1 Mandarin and Korean pretest scores by vowel (percent correct out of 5 for each vowel)

|       | æ     | ɛ     | iː    | ɪ     | u     | ʊ     |
|-------|-------|-------|-------|-------|-------|-------|
| M-Pre | 66.04 | 59.62 | 86.04 | 81.88 | 73.58 | 63.02 |
| K-Pre | 70.58 | 60.58 | 71.76 | 66.48 | 64.12 | 55.3  |

According to the Speech Learning Model (SLM), Mandarin speakers should have the greatest ease identifying sounds that are either *identical* to a same-labeled vowel in their L1, or completely *new*. Vowels that have the same label but are not the same as the L1 counterpart are predicted to be difficult to perceive and produce. This latter category subsumes the perception and production of a sound that is allophonic variant in the L1 and contrastive in the L2. Table 6-2 summarizes predicted difficulties, according to a traditional interpretation of SLM.

Table 6-2. SLM – traditional – predicted difficulties.

| SLM - traditional | æ | ɛ | i | ɪ | u | ʊ |
|-------------------|----------|----------------|-------------------|----------|-------------------|----------|
| Mandarin | new-easy | similar - hard | identicial- easy | new-easy | identicial- easy | new-easy |
| Korean | new-easy | similar - hard | identicial- easy | new-easy | identicial- easy | new-easy |

Traditional SLM accurately predicted that /i/, /ɪ/ and /u/ would be easy for the Mandarin speakers and that /ɛ/ would be difficult. SLM failed to predict that Mandarin speakers would have difficulty with /ʊ/ and /æ/. SLM correctly predicted that Korean speakers would have an advantage identifying /i/ and /æ/ and that they would have difficulty with /ɛ/. SLM failed to anticipate Korean speakers' difficulty with /ɪ/, /u/, and /ʊ/.

Adapting SLM to the vowel similarity data from Wu (2011), Thomson (2009) and Yang

(1996), /æ/ /ɛ/ /u/ and /ɪ/ are the least similar to any vowel in Mandarin. They would therefore, be considered new and easy. /ʊ/ and /i/ would be identical to /ɣ/ and /i/ vowels in Mandarin and therefore easy to perceive and identify. For Korean speakers, /æ/ /u/ /ɪ/ and /ʊ/ would be "new" and therefore easy. Given the indeterminate status of /ɛ/ in Korean, it is unclear how it should fare. For Korean speakers, /i/ would be similar – but not identical – and therefore difficult. Table 6-3 provides a summary of the adapted SLM predictions.

Table 6-3 adapted SLM

| Wu,Thomson, Yu | æ | ɛ | i | ɪ | u | ʊ |
|---|---|---|---|---|---|---|
| Mandarin | new-easy | new-easy | identicial - easy | new-easy | new-easy | identicial - easy |
| Korean | new-easy | ? | similar - hard | new-easy | new-easy | new-easy |

Adapted SLM accurately predicted that /i/, /ɪ/ and /u/ would be easy for the Mandarin speakers, but failed to predict relative difficulty with /æ/ /ɛ/ and /ʊ/. Korean speakers were predicted to have difficulty with /i/ since it was similar but not identical to the L1 /i/; this was not the case, as it was the most accurately identified vowel for the Korean speakers in the pretest. Adapted SLM correctly predicted that Korean speakers would have an advantage identifying /æ/ but failed to predict their difficulty with /u/ /ɪ/ and /ʊ/. The status of /ɛ/ made it difficult to make any predictions, but Korean speakers clearly had difficulty accurately identifying this sound.

In the pretest on nonce words, ASP does not offer particular predictions since these are not lexical items nor are they being presented with two sounds to compare. However, given the similarity between Mandarin and NAE /i/ ASP may have predicted that this would have been the easiest sound for Mandarin speakers – which was born out in pretest scores. Yang (1996) found that Korean /i/ and NAE /i/ were relatively similar and consistent with this, Korean speakers identified /i/ with the highest accuracy rates followed by /ae/. In general, the Korean speakers performed in a manner consistent with previous research that found difficulty with tense-lax vowel contrasts with a distinct advantage at identifying tense vowels. Both groups of speakers, Mandarin and Korean, identified tense vowels accurately more often than their lax counterparts.

In general, neither the predictions based on a traditional instantiation of SLM – nor the adapted model seemed to make predictions that fit the data.

**6.3 Training**

Although there were differences in accuracy rates on pretest scores, no significant differences were found between the amount of training – measured by the number of sound files played - between Mandarin and Korean participants. Given the differences in the protocols instantiated in each paradigm, within group differences in the amount of training were examined. Only Korean speakers evinced a significant difference, training less in Paradigm D than any other paradigm. Recall that the Korean students began the training with lower scores in pretests than Mandarin participants. It is possible that limiting variation in training material may render training more tedious than helpful and without the incentive of payment and the control of a laboratory setting, students – perhaps particularly those who have more difficulty identifying the

target material – are more likely to abandon the effort. In other words, rank boredom may affect the amount of training participants are willing to complete.

## 6.4 Post-tests: Real Words

On the test of real word identification, Mandarin speakers as a group made significant gains; Korean speakers taken in *toto* did not. However, considering differences by paradigm, both Mandarin and Korean speakers in Paradigm A, consisting of many speakers and three speech rates, and B, many speakers and one speech rate, showed significantly greater accuracy in identifying real words in the post-test. Those participants who trained in Paradigm C, one speaker and three speech rates, and D, one speaker, one speech rate, did not improve significantly. In terms of transferring learning from nonce words and strictly phonetic identification to real word identification, it would appear that multiple speakers benefited both Mandarin and Korean groups. For those participants in Paradigms C and D who made no significant gains, there is an increasing body of evidence that orthographic representations, phonological encoding and memory for phonologic forms interact in numerous ways – and these interactions will differ depending on the writing system in question (Bassetti, 2006; Cutler, 2012; Escudero et al., 2008; Hayes-Harb et al., 2010; Showalter et al., 2013).When new words are learned, they are learned as an orthographic form tied to a phonological form. For example, Borden et al. (1983) found that L2 speakers engaged in a production task were better at imitating nonsense syllables than real words. These researchers suggest that the presentation of real words activated previously stored, inaccurate phonetic representations (Thomas, 2010). Similarly, perhaps, the Korean listeners in this study, who had more difficulty in pretest conditions, were activating a previously stored, but inaccurate phonological forms such that the phonetically

relevant information at hand was bypassed. Training with only one voice, as those Mandarin and Korean participants in Paradigms C and D, did not provide adequate variation to impact previously stored representations of real words.

## 6.5 Post-tests: Nonce Words

Both Mandarin and Korean participants made significant gains between pretest and post-test scores on nonce words. Considering scores by training paradigm, all groups made significant gains in identifying vowels in nonce words, save for Korean speakers in training paradigm D, the least variable condition and the condition within which Korean participants trained the least. As such, it is not possible to say whether lack of improvement was due to the interaction of L1, Korean, and Paradigm D or whether it was an artifact of the amount of training. The summary table of scores by L1 and Paradigm are re-presented in Table 6-4 for convenience.

Table 6-4 Average Percent accurate and change: Scores by L1 and Paradigm

|  | Paradigm | %PreRW | %PostRW | %Change | %PreNonc | PostNonc | %Change |
|---|---|---|---|---|---|---|---|
| Mandarin | A | 71.98 | **77.47*** | 5.49 | 68.1 | **83.57*** | 15.47 |
| Mandarin | B | 73.26 | **77.51*** | 4.25 | 74.05 | **84.76*** | 10.71 |
| Mandarin | C | 69.63 | 68.44 | -1.18 | 71.39 | **79.74*** | 8.35 |
| Mandarin | D | 68.16 | 70.73 | -2.56 | 70.3 | **79.72*** | 9.42 |
|  |  |  |  |  |  |  |  |
| Korean | A | 56.09 | **64.10*** | 8.01 | 67.92 | **84.71*** | 16.79 |
| Korean | B | 58.33 | **62.56*** | 4.23 | 65.83 | **75.42*** | 9.59 |
| Korean | C | 64.36 | 63 | -1.36 | 65.33 | **76.33*** | 11 |
| Korean | D | 68.27 | 66.67 | -1.6 | 65.42 | 71.25 | 5.83 |
|  |  |  |  |  |  |  |  |
| Mandarin | Control | 67.63 | 65.06 | -2.56 | 70 | 75.42 | 5.42 |
| Korean | Control | 64.47 | 65.57 | 1.1 | 69.44 | 78.89 | 3.81 |

Significant gains between pre- and post-test scores in identifying vowels in nonce words were in evidence across all paradigms (save for the Korean learners in paradigm D). Given these results, it is not clear that any particular kind of variation is more helpful than another or even a lack of variation. The next section considers improvement by vowel.

**6.6 Improvement by Vowel: Mandarin**

Mandarin speakers improved significantly on their perception of every vowel except /i/ which may be due to ceiling effects; /i/ was the most accurately identified vowel in the pretest, identified with 86% accuracy reaching 90.19% accuracy in post-test. Close behind /i/ in terms of accuracy on pretest were /ɪ/ and /u/ although these two also made significant improvement between pre- and post-test. The greatest improvement for the Mandarin speakers was on /ʊ/ which reached 86.42% accuracy compared to 63.02% in pretest. The Mandarin group had the most difficulty with /ɛ/ and /ae/ in the pretest and although they made significant improvements, /ɛ/ only reached 70.19% accuracy in post-test followed closely by its contrasting partner /ae/ which reached 74.72% accuracy. Accuracy rates in pre-and post-test by vowel for the Mandarin group are repeated in Table 6-5.

Table 6-5 (repeated from 5-15) Mandarin gains by vowel (N=5)

|          | æ      | ɛ       | i     | ɪ      | u      | ʊ       |
|----------|--------|---------|-------|--------|--------|---------|
| M-Pre    | 66.04  | 59.62   | 86.04 | 81.88  | 73.58  | 63.02   |
| M-Post   | 74.72  | 70.19   | 90.19 | 88.68  | 81.89  | 86.42   |
| % change | **8.68***  | **10.57***  | 4.15  | **6.80***  | **8.31***  | **23.40***  |

The Automatic Selective Perception model (ASP) predicted that perception of vowels in such a training paradigm should be relatively easy since the learners are listening in a particular way, listening for phonetic details. In this mode, perception of all the contrasts should be relatively easy – although there was some question regarding /ɛ/ given its status as a possible allophone in Mandarin. Note that this phone was the most difficult in the pretest and while the gains were significant, after training, it still only reached 70% accuracy rate. Similarly, /ae/ only reached 74.72% accuracy in post-test. A Table with predictions according to the ASP model is presented in Table 6-6.

Table 6-6. ASP predictions

| ASP - Mandarin | ae | ɛ | iː | ɪ | u | ʊ |
|---|---|---|---|---|---|---|
| phonological perception | hard | hard | easy | hard | easy | hard |
| phonetic perception | easy | ? | easy | easy | easy | easy |

While ASP asserts that all contrasts will be equally easy given the simplicity of material, mode of listening and the focus of attention, clearly some sounds are more difficulty for Mandarin speakers to disambiguate than others. ASP allows for the likelihood that listeners will be sensitive to context specific variation (due to coda, speech rate, voice) but a mechanism for predicting those effects are not yet part of the model.

### 6.7 Improvement by Vowel: Korean

Korean speakers made gains in the accurate identification of all vowels save /ae/ and /i/, which were the two most accurately identified vowels in the pretest. That being said, both vowels were only around 70% accurate in pretest with /ae/ making no progress in post-test and /i/

nominally so. Korean participants improved the most on /ʊ/ attaining 78.82% accurate identification. Although the distribution of acuity after training similar to that of the Mandarin speakers, that is /i/~/ɪ/ followed by /u/~/ʊ/ followed by /ae/~/ɛ/, overall scores are lower for the Korean speakers. Table 6-7 re-presents Korean gains by vowel.

Table 6-7 (repeat from table 5.16) Korean gains by vowel

|  | æ | ɛ | i | ɪ | u | ʊ |
|---|---|---|---|---|---|---|
| K-Pre | 70.58 | 60.58 | 71.76 | 66.48 | 64.12 | 55.3 |
| K-Post | 70.58 | 74.12 | 78.82 | 81.76 | 76.48 | 78.82 |
| % change | 0 | **13.54*** | 7.06 | **15.28*** | **12.36*** | **23.52*** |

ASP predicts that perception of vowels in a training paradigm study such as this should be relatively easy since the learners are listening specifically for phonetic details. Given that the participants are learning to perceive a contrast, listening in this mode of perception, should make the differences obvious and easy to perceive. However, note that Korean participants failed to make any gains in their perception of /ae/, no significant gains in /i/ and that overall scores remained relatively low. Table 6-8 summarizes ASP predictions based on perceptual mode and vowel.

Table 6-8 Korean: ASP predictions

| ASP - Korean | ae | ɛ | iː | ɪ | u | ʊ |
|---|---|---|---|---|---|---|
| phonological perception | hard | easy | easy | hard | easy | hard |
| phonetic perception | easy | easy | easy | easy | easy | easy |

It's possible that if the Korean participants needed more training given that they were less able to identify certain vowel contrasts in the pretests.

The next section considers changes in the perception of individual vowels by the paradigm within which the groups trained.

**6.8 Vowel by Paradigm**

Mandarin speakers did not differ on the amount of training by vowel or by paradigm. Different gains between individual vowels then suggests a result of the paradigm within which a participant trained. For the Mandarin speakers, /ʊ/ improved under all training conditions. Given the similarity between Mandarin /ɣ/ and /ʊ/, perhaps more or less variation was irrelevant for improvement. Mandarin speakers were also already quite strong at identifying /u/ in the pretest, despite its significant phonetic differences from NAE /u/. Tense vowels, /æ/, /i/,/u/ all improved under Paradigm A, with multiple speakers at three speech rates. Of the two remaining lax vowels, /ɛ/ and /ɪ/, Mandarin speakers improved significantly identifying /ɛ/ only in Paradigm B, multiple speakers at one speech rate; identification of /ɪ/ only improved in Paradigm D, one speaker at one speech rate. It is of note that both lax vowels made significant improvements in paradigms with no speech rate changes.

Table 6-9 (repeated from 5-17) Mandarin: average % gain scores by vowel by paradigm

| Vowel | Paradigm | | | |
|---|---|---|---|---|
| | A | B | C | D |
| æ | **18.57*** | 1.43 | 6.15 | 8.33 |
| ɛ | -2.86 | **25.71*** | 12.31 | 6.67 |
| i | **14.29*** | 2.86 | 4.62 | -6.67 |
| ɪ | 11.43 | 2.86 | 1.54 | **11.67*** |
| u | **17.14*** | 12.86 | 7.69 | -6.67 |
| ʊ | **20*** | **17.14*** | **27.69*** | **30.00*** |

Korean speakers did not differ on the amount of training by vowel, nor were there differences in the amount of training between those in Paradigm A, B, and C, although those Korean students in Paradigm D trained significantly less overall. For Korean participants, Paradigm C, consisting of a single speaker, with increasing speech rates, was clearly the best training condition for the majority of vowels (Table 6-10).

Table 6-10 (repeated from Table 5-19) Korean: average % gain scores by vowel by paradigm

| Vowel | Paradigm | | | |
|---|---|---|---|---|
| | A | B | C | D |
| æ | 10 | 10 | 0 | -20 |
| ɛ | **22.50*** | 7.5 | **16.00*** | 7.5 |
| i | 10 | 5 | **14.00*** | -2.5 |
| ɪ | 17.5 | 17.5 | **14.00*** | 12.5 |
| u | 7.5 | 0 | **22.00*** | 17.5 |
| ʊ | 25 | 17.5 | **28.00*** | 22.5 |

Korean students had difficulty identifying /ɛ/ in the pretest but made significant gains in Paradigm A and C; these two training paradigms share a common feature: rate variation. /æ/ which was relatively well-identified in pretest, made no gains in any paradigm.

Different kinds of variation interact with different vowels and different L1s to create a more complex picture than HVPT might suggest. Mandarin speakers' most modest gain was a 14 perent gain in identifying /i/ in Paradigm A; compare this to a 30% gain for /ʊ/ in Paradigm D. Korean speakers made the most modest gains, 14%, identifying /i/ and /ɪ/ in Paradigm C; their greatest gain was in identifying /ʊ/ also in Paradigm C. For the Korean speakers, Paradigm C provided the best learning condition. Mandarin speakers improved more variously, but with an affinility multiple speakers were of greater benefit than rate changes. For the Korean group, changes in rate may be illustrative of co-articulation effects in vowels that allowed them to generalized to tokens produced by different speakers.

The next section addressess the impact of the coda on perceptual acumen of the trainee participants.

**6.9 Coda Condition**

There is clear evidence that the coda condition differentially affects the accuracy of vowel identification. For example, for the Mandarin speakers, the nasal coda /n/ has the greatest detrimental effect on lax vowels in general. The nasal coda /m/ however, for both Mandarin and Korean speakers has a greater impact on /æ/~/ɛ/. For the Korean participants, nasal codas represented 62.78% of errors identifying /ae/ and /ɛ/ (recall that /ae/ identification did not improve in any paradigm for the Korean speakers).

Mandarin and Korean, it is surprising to note the differential effects. A more general explanation for the deleterious effects that nasal codas have on accurate vowel identification may be that co-articulation patterns in English differ from Mandarin (and Korean) patterns of co-articulation. For example, Solé (1992) examined differences in anticipatory nasalization of

vowels between Spanish and English speakers and found language specific patterns of anticipatory vowel nasalization that differed both in terms of onset and rate-o*f*-speech effects. Li (2008) in an acoustic study comparing Mandarin and English speakers' production of vowel-nasal (VN) sequences found that Mandarin VN sequences were more strongly co-articulated than English productions of VN – and that English VN and VŊ had significant length differences. Given that nasalization results in an overall dampening of the formant structure of the preceding vowel, it seems plausible that even small differences in co-articulation effects could impact vowel identification.

Finally, although Korean allows voiceless consonants in codas, they account for just over 53% of the errors for NAE /i/ - a vowel that is very similar to the Korean /i/. In Mandarin, both high tense vowels, /i/ and /u/ are disproportionately erroneously identified when /s/ is in the coda; all three tense vowels, /ae/, /i/ and /u/ are more prone to error when a /p/ is in the coda, but not /b/. To date, much of the research on vowel identification has used steady-state, bare vowels (Nearey & Assmann, 1986) or CV training tokens (Mok, 2012; Thomson, 2009; Wu, 2011), others have used a very limited set of real word tokens manipulated to exaggerate the relevant features. Given that one of the persistent questions regarding vowel identification and classification is the contribution of adjacent segments and in light of the effect that codas seemed to have in this training study, consonant context and co-articulation effects are areas that warrant further research (e.g., see Connine & Darnieder, 2009). They must also be taken into consideration in future vowel training studies and in L2 teaching practices.

In considering RQ 1, "What kinds of variation results in the most robust vowel categories?" Clearly there is no one answer. Korean speakers, perhaps because they were lower proficiency, benefited most with variation in rate. One of the trained vowels /ae/ made no

improvement in any training paradigm. Mandarin speakers benefited most with variation in speaker and voice. Except that /ɪ/ made gains in Paradigms D – one voice, at one speech rate – and /ʊ/ improved under every condition. In terms of variation in coda, there was evidence that vowel identification could be facilitated or inhibited depending on which vowel and which coda.

In sum, the results from this work indicate that laboratory methods can be successfully employed in the classroom. It also demonstrates that predicting which vowels in the L1 will be more or less difficult will depend upon context – both in terms of the training protocols as well as the consonant conditions. It is also evident that input with acoustic variability due to speaker voice, speech rate, or consonantal context differentially impacts the course of novel vowel category formation and that some vowels are more difficult to learn than others. In general, however, it is borne out that training that is focused on difficult sounds helps in ways that general input does not.

## Chapter 7: Conclusion

This work reports on the outcomes of training L2 English learners to perceive the English tense-lax vowel contrast outside the context of the laboratory, as part of course work in an upper level ESL pronunciation class and as such, under conditions of autonomy. Training consisted of naturally produced, unenhanced nonce words designed to have a representative distribution of consonant conditions. Training tokens were distributed between four training paradigms designed to constitute different degrees of variation. A total of 87 graduate students, 53 L1-Mandarin and 34 L1-Korean speakers, completed training as homework over the course of six days.

The Mandarin student participants in paradigm A, B, C, D made significant gains in their ability to identify tense-lax vowel contrasts in nonce words. Only Mandarin speakers in Paradigms A and B, however, improved on real word identification in post-test. Korean participants in paradigm A, B and C made significant improvements in their ability to identify tense-lax vowel contrasts in nonce words but, like their Mandarin counterparts, only those in Paradigm A and B made significant gains on real word post-tests.

There is evidence that some learners benefited from more constrained variation in the training material, e.g., limited to rate change rather than multiple speakers. And yet only those who trained on multiple speakers made gains in real word identification. It may be the case that less proficient learners benefit from less variation but that as they gain proficiency, increasingly variable exemplars of target sounds should be introduced into their training.

Closer analysis revealed that some vowels were more difficult to master than others, e.g. /ae/ and /ɛ/ were difficult for both groups in both pre-and post-test. All groups were better at identifying tense vowels versus lax vowels and this strength was maintained in post-test scores.

Consonant context also had a significant effect on accurate vowel perception, albeit differentially. For example, /ɛ/ with a nasal coda was misidentified at more than twice the rate of /æ/ followed by a nasal – the opposite pattern was found for coda /g/ which had a deleterious effect on /æ/ identification at five times the rate of /ɛ/. Given that spectral changes (changes in formant frequency) between consonants and vowels contain information necessary for accurate vowel identification (Assman & Katz, 2005; Hillenbrand et al., 2001) and since patterns of spectral change can vary by dialect (Jacewicz & Fox, 2013), then it is not unreasonable to assume that they would also vary by language. As such, consonant-to-vowel co-articulation should be included as a variable in any future vowel training studies.

An additional finding is the (lack of) impact of training on the extant lexicon for those who trained in Paradigm C and D. While participants made gains in their ability to perceive the vowel contrasts under a variety of training paradigms and in varied consonant contrasts, the same ability did not generalize to previously encoded lexical items. There may be an implication for when, in the progression of L2 learning, such perceptual training might be employed. For example, perceptual training on difficult phonemes should begin early – perhaps with a training paradigm modeled after Paradigm C in this work, where the less proficient speakers made the greatest gains. Training should begin prior to the establishment of a large lexicon. As learners become more proficient, increasing the amount of variation, such as that in training Paradigm A, may provide more facilitation in vowel learning.

However one were to implement phoneme training into language teaching, there seems to be sufficient evidence that this kind of training and these kinds of experimental methods can successfully be employed in a classroom environment.

## 7.1 Teaching Pronunciation

I began this dissertation with a quote from Liberman's (2008) online blog in which he questions why something as effective and efficient as HVPT has made almost no inroads into educational settings. Although the experimental evidence for this dissertation was collected in such a context, efforts to make it available to other ESL pronunciation classes on location were certainly not met by instructors with universal enthusiasm. I can only speculate that the mantra of authentic communication in meaningful contexts with a concomitant body of scholarly literature focused on communicative competence has left little room for such methods. Given such a pedagogical orientation, perceptual training that targets such relative minutiae and does it without meaningful context, is, perhaps, viewed with skepticism. However, I reiterate that this same insistence on meaningful communication is exactly the context that inhibits the perception and subsequent acquisition of novel phonemes. Haslam (2011) addressed this issue specifically in her dissertation wherein she compared the effects of typical laboratory-based perceptual training with laboratory training that required "lexical access and interpretation of linguistic context," which she hypothesized was more like 'real world learning.' She found that those listeners trained in the 'real world' paradigm, were much less accurate in perception post training than those trained in the 'typical laboratory' way.

Resistance to implementing perceptual training takes place within an area of ESL teaching that is already relatively marginalized. Part of this marginalization is based on the

misapprehension that pronunciation instruction has little or no effect on learners' productions; if the target is native-like pronunciation, than perhaps a certain amount of skepticism is warranted. However, in an ever more sophisticated and interconnected world, few instructors - and frankly, fewer learners - desire such a target. Increasingly, the emphasis is on clear, comprehensible speech, a goal that could be assisted through phoneme training designed to circumvent the task complexity of extracting meaning from real speech, in real time.

## References

Ames, H., & Grossberg, S. (2008). Speaker normalization using cortical strip maps: a neural model for steady-state vowel categorization. *The Journal of the Acoustical Society of America*, *124*(6), 3918-3936.

Anderson, J. R. (2000). *Cognitive psychology and its implications* (5th ed.). New York: Worth Publishers.

Assmann, P.F. and Katz, W.F. (2005). Synthesis fidelity and vowel identification. *Journal of the Acoustical Society of America, 117(2*), 886-895.

Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford University Press.

Barlow JA, & Gierut JA, (2003). Minimal pair approaches to phonological remediation. *Seminars in Speech and Language* (23), 57–67.

Bassetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language and Literacy, 9*, 95–114.

Becker, W. E. (2001). What does the quantitative research literature really show about teaching methods? Paper presented at the *Scholarship of Teaching and Learning Colloquium*, Indiana University, Bloomington, IN.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.

Best, C. T., McRoberts, G., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America,* (*109*), 775-794.

Best, C. T., & Tyler, M. D. (2011). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34). Amsterdam, Netherlands: John Benjamins.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41-68). Cambridge: Cambridge University Press.

Bod, R., & Cochran, D. (Eds.). (2007). Exemplar-based models of language acquisition and use. *Proceedings of the 19<sup>th</sup> European Summer School in Logic, Language and Information (ESSLLI) 2007*, Dublin, Ireland.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer /Computer program/. Version 5.3.72, retrieved 17 April 2014 from http://www.praat.org/

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In P. Avery, E. Dresher, & K. Rice (Eds.), *Phonological Contrast* (pp. 143-172). Berlin, Germany: Mouton.

Borden, Gerber, and Milsark (1983). Production and perception of the /r/-/l/ contrast in Korean adults learning English. *Language Learning*, 33, 499-526.

Bowen, C. (1998). Give me five: A broad-based approach to phonological therapy. *Proceedings of the New Zealand Speech-Language Therapists' Association Biennial Conference*, Dunedin, New Zealand.

Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In Hansen Edwards, J. G. and Zampini, M.L. (Eds.), *Phonology and Second Language Acquisition*, (pp. 287-308), John Benjamins Publishing Company.

Bradlow, A. R. and Bent, T. (2008) Perceptual adaptation to non-native speech. *Cognition*, *106* (2), 707–729.

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

Brauna, B., Kochanski, G., Grabe, E., and Rosner, B.S. (2006). Evidence for attractors in English intonation. *Journal of the Acoustical Society of America,* 119 (6), pp. 4006-4015

Brown, D. J., & Proulx, M. J. (2013). Increased signal complexity improves the breadth of generalization in auditory perceptual learning, *Neural Plasticity,* vol. 2013, Article ID 879047.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change, 14*, 261–290.

Carroll, S. E. (1999). Putting "input" in its proper place. *Second Language Research*, *15*(4), 337–388.

Carroll, S. E. (2001). *Input and evidence: The raw material of second language acquisition.* Amsterdam: John Benjamins.

Carroll, S. E. (2004). Segmentation: learning how to "hear words" in the L2 speech stream. *Transactions of the Philological Society*, *102*(2), 227–254.

Chang, Y-H. S. (2010). Training Mandarin speakers to identify English vowels. Paper presented at *American Association for Applied Linguistics (AAAL)*, Atlanta, GA.

Chen, Y. (2006). Production of tense-lax contrast by Mandarin speakers of English. *Folia phoniatrica et logopaedica, 58*(4), 240-249.

Cheng, C. C. (1973) *A Synchronic Phonology of Mandarin Chinese*, Mouton, The Hague.

Cheng, R. L. (1966). Mandarin Phonological Structure, *Journal of Linguistics* 2(2), 135–262.

Cole, J., & Shattuck-Hufnagel, S. (2011). The Phonology and Phonetics of Perceived Prosody: What do Listeners Imitate? In *INTERSPEECH* (pp. 969–972). ISCA.

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review, 11*, 1084–1089.

Connine, C. M., & Darnieder, L. M. (2009). Perceptual learning of co-articulation in speech. *Journal of Memory and Language, 61*(3), 412-422.

Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics, 5*, 161-170.

Cornell, S. A., Lahiri, A., & Eulitz, C. (2011). "What you encode is not necessarily what you store": Evidence for sparse feature representations from mismatch negativity. *Brain research, 1394*, 79-89.

Cousin, E., Peyrin, C., Pichat, C., Lamalle, L., Le Bas, J. F., & Baciu, M. (2007). Functional MRI approach for assessing hemispheric predominance of regions activated by a phonological and a semantic task. *European Journal of Radiology, 63*(2), 274-85.

Cutler, A., & Davis, C. (2012). An orthographic effect in phoneme processing, and its limitations. *Frontiers in Psychology*, 3, 18.

De Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics*, 28(4), 441–465.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499–533

DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42-63). New York: Cambridge University Press.

Dupoux, E., Pallier, C., Sebastián, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language, 36,* 406–421.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*(3), 379-397.

Elbert, M., Powell, T. W., & Swartzlander, P. (1991). Toward a technology of generalization: How many exemplars are sufficient? *Journal of Speech and Hearing Research, 34*, 81–87.

Ellis, N. C. (2004). The processes of second language acquisition. *Form-meaning Connections in Second Language Acquisition*, 49–76. Mahwah, NJ: Erlbaum.

Ellis, R. (1998). Teaching and research: options in grammar teaching. *TESOL Quarterly*, 32 (1), 39-60

Ellis, R. (2004). Definition and Measurement of L2 Explicit Knowledge, *Language Learning 54*, pp.227-275.

Ellis, R. (2005). Principles of instructed language learning. *System 33* pp. 209–224

Ellis , R. (2006). Current Issues in the Teaching of Grammar: An SLA Perspective . *TESOL QUARTERLY Vol. 40* (1), 83-107.

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

Escudero, P., Hayes-Harb R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics, 36*, pp. 345–60.

Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech, 53*, pp. 343–65.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47-65.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop conso- nants produced in a second language. *Journal of the Acoustical Society of America, 8*9, 395–411.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26*, pp. 1-34.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience* , pp. 233-277. Timonium, MD: York Press.

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

Flege, J., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, *99*, 1161-1173.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 110-125.

Gardner, R. (1985). *Social psychology and second language learning: The role of attitude and motivation*. London, UK: Edward Arnold.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27,* 18121

Gass, S. (1988). Integrating research areas: a framework for second language studies. *Applied Linguistics 9,* 198–217.

Gass, S. (1997). *Input, interaction, and the second language learner.* Mahwah NJ: Erlbaum.

Goldinger, S. D. (1998) Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–79.

Goldinger, S. D. (1996). Words and voices: Episodic traced in spoken word identification. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *22*(5), 1166-1183.

Goldinger, S. D. (1997). Words and voices: Perception and production in an episodic lexicon. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 33-66). San Diego, CA: Academic Press.

Goldinger, S.D., Pisoni, D.B., & Logan, J.S. (1991). On the nature of talker variability effects on serial recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 152-162.

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds: Reallocation of neural substrates. *NeuroImage, 21*, 494–506.

Gordon, P.C., Eberhardt, J.L., & Rueckl, J.G., (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology, 25(1)*,1-42.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r".

    *Neuropsychologia, 9,* 317-323

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and

    ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3),

    311–343.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In

    O.-S. Bohn & M. J. Monro (Eds.), *Language experience in second language speech*

    *learning: In honor of James Emil Flege* (pp. 58-77). Amsterdam, Netherlands: John

    Benjamins.

Ha, S., Johnson, C. J, & Kuehn, D. P. (2009). Characteristics of Korean phonology: Review,

    tutorial, and case studies of Korean children speaking English. *Journal of*

    *Communication Disorders, 42*(3), 163–179

Harnsberger, J. D. (2001). On the relationship between identification and discrimination on non-

    native nasal consonants. *Journal of the* Acoustical Society of America, 110, 489-503.

Haslam, M. (2011). The effect of perceptual training including required lexical access and

    meaningful linguistic context on second language phonology. Unpublished doctoral

    dissertation, University of Utah, Salt Lake City, UT. Retrieved from

    http://content.lib.utah.edu/utils/getfile/collection/etd3/id/416/filename/image

Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words:

    Effects of orthographic and auditory input. *Language and Speech, 53*, 367–81.

Hazan, V. & Barrett, S. (2000). The development of phonemic categorization in children aged 6–

    12. *Journal of Phonetics*, 28, 377–396.

Hickok & Poeppel, (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition 92,* 67–99

Hillenbrand and Gayvert, (1993). Vowel Classification Based on Fundamental Frequency and Formant Frequencies. *Journal of Speech, Language, and Hearing Research Vol.36*, 694-700

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America, 109*(2), 748-763.

Hodson, B. W., & Paden, E. P. (1991). *Targeting intelligible speech: A phonological approach to remediation*. Pró-ed Austin, TX. Retrieved from http://www.getcited.org/pub/102865022

Iverson, P. & Evans, B.G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *Journal of the Acoustical Society of America, 126*(2), 866-77.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., & Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, B47-B57.

Jacewicz, E., and Fox, R. A. (2013). The effects of dialect variation on speech intelligibility in a multitalker background, *Applied Psycholinguistics*, pp 1-18, available on CJO2014. doi:10.1017/S0142716413000489.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ae/–/eh contrast by Francophones. *Perception and Psychophysics 40*: 205–215.

Jamieson, D., & Morosan, D. (1989). Training new, nonnative speech contrasts: A comparison prototype and perceptual fading techniques. *Canadian Journal of Psychology, 43*, 88–96.

Jenkins, J. Strange, W. & Edman, T.R. (1983). Identification of vowels in "vowelless" syllables. *Perception & Psychophysics, 34*, 441-450.

Johnson, K. (2004) Cross-linguistic perceptual differences emerge from the lexicon. In Augustine Agwuele, Willis Warren, and Sang-Hoon Park (Eds.), *Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in speech production and perception* (pp. 26-41). Sommerville, MA: Cascadilla Press.

Johnson, K. (1997). Speech perception without speaker normalization: an exemplar model. In K. Johnson & J. Mullenix (Eds.), *Talker Variability in Speech Processing* (pp. 145-166). Salt Lake City, UT: Academic Press.

Jongman, A., and Wade, T. (2007). Acoustic variability and perceptual learning: The case of non-native accented speech. In O-.S. Bohn and M. Munro (Eds.), *Language experience in second language speech learning.* Amsterdam, Netherlands: John Benjamins.

Kang, S.-K. (1996). Acoustic analysis of South and North Korean vowels. *Language Research, 32*, 1–18

Kazanina, N., Philips, C., & Idsardi, W. (2006, July 25). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences, 103*(30), 11381-11386. Retrieved from http://www.pnas.org/content/103/30/11381.long

Kewley-Port, D., Bohn, O.-S., & Nishi, K. (2005). The influence of different native language systems on vowel discrimination and identification. *Journal of the Acoustical Society of America, 117*, 2399.

Kingston, J. (2003). Learning foreign vowels. *Language and Speech*, *46*(Pt 2-3), 295-349.

Kondaurova, M. V., & Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: comparison of three training methods. *Journal of Phonetics, 38*(4), 569-587.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford, UK:Pergamon Press.

Krashen, S. (1985). *The input hypothesis: Issues and implicatio*ns . New York: Longman.

Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T. Kiritani, S. and Iverson, P (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months, *Developmental Science 9:2*, F13–F21.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. deBoysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259-274). Dordrecht, Netherlands: Kluwer Academic Publishers.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the perceptual magnet effect. In W. Strange (Ed.), *Speech Perception and Linguistic Experience* (pp. 121-154). Timonium, MD: York Press.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606–608.

Larson-Hall, J. (2006). What does more time buy you? Another look at the effects of long-term residence on production accuracy of English /r / and / l / by Japanese speakers. *Language and Speech 49*(4), 521-548.

Li, T., & Fu, Q-J. (2007). Perceptual adaptation to spectrally shifted vowels: Training with nonlexical labels. *Journal of the Association for Research in Otolaryngology, 8*(1), 32-41.

Li, Ya (2008). *Mandarin Speakers Production of English and Mandarin Post-Vocalic Nasals: An Acoustic Approach*. Unpublished dissertation, University of Victoria, Victoria, BC, CA.

Liberman, M. (2008). HVPT /blog: filed under Language teaching and learning/. Retrieved from http://languagelog.ldc.upenn.edu/nll/?p=328

Lin, Y. H. (1989) Autosegmental Treatment of Segmental Processes in Chinese Phonology, PhD dissertation, University of Texas, Austin.

Lin, Y. H. (1992). Vocalic Underspecification in Two Mandarin Dialects: A Case Against Radical Underspecification. *Journal of East Asian Linguistics 1(3)*, 219–253.

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

Lively, Pisoni, Yamada, Tokhura, & Yamada, (1994). Training Japanese listeners to identify English /r/ and /l/ Long term retention of new phonetic categories *Journal of the Acoustical Society of America,* 96, 2076-2087

Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 351–377). Timonium, MD: York Press.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, 109, 376–400.

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

Long, M. H. (1985). Input and second language acquisition theory. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 377-393). Rowley, MA: Newbury House.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W.C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego, CA: Academic Press.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty& J. Williams (Eds.) *Focus on form in classroom second language acquisition* (pp. 197-261). New York: Cambridge University Pres

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

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance, 33*(2), 391–409.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*(5417), 522–523.

Martin, C.S., Mullennix, J.W., Pisoni, D.B., & Summers, W. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 676-684.

McAllister, R. (2001). Experience as a factor in L2 phonological acquisition. *Working Papers, 49* (pp. 116-119), Department of Linguistics, Lund University, Sweden.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the r-l contrast to Japanese adults: Predictions of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, and Behavioral Neuroscience*, *2*(2), 89-108.

McClaskey, C. L., Pisoni, D. B. and Carrell, T. D. (1983). Effects of transfer of training on identification of a new linguistic contrast. *Perception & Psychophysics*, *34*, 323-330

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science, 317*, p. 82.

Mok, P. (2012), Does vowel inventory density affect vowel-to-vowel coarticulation? *Language and Speech 56*(2), 191–209.

Mullennix, J. W., and Pisoni, D. B. (1990). "Stimulus variability and processing dependencies in speech perception," *Perception and Psychophysics, 47(4)*,379–390.

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology, 21*, 422–432.

Nearey, T.M. and Assmann, P.F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America, 80*, 1297-1308

Nguyen, N. (2011). Representations of speech sound patterns in the speaker's brain: Insights from perception. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 359-368). Oxford, NY: Oxford University Press.

Nguyen, N., Wauquier, S., & Tuller, B. (2009). The dynamical approach to speech perception: from fine phonetic detail to abstract phonological categories. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to phonological complexity* (pp. 193–217). Berlin, Germany: Mouton de Gruyter.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*, 62–85.

Nittrouer, S. (2005). Perception of steady-state vowels and vowelless syllables by adults and children. *Journal of the Acoustical Society of America, 117*(4), 2402-2402.

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

Norbre-Oliveira, (2007), The Acquisition of Second Language Speech (/i-ih/, /ae-eh/, and /uh-u/) by Brazilian EFL students. *New Sounds 2007: Proceedings of the Fifth International Symposium on The Acquisition of Second Language Speech*.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.

Nosofsky, R. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39- 57.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*(1), 42–46.

Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics, 57,* 989-1001.

Nygaard, L. C., and Pisoni, D. B. (1998). Talker-specific learning in speech perception, *Perception & Psychophysics, 60(3)*, 355–376.

Nygaard, L. (2005). Perceptual integration of linguistic and nonlinguistic properties of speech. In D. Pisoni & R. Remez (Eds.), *Handbook of speech perception* (pp. 390–414). Oxford, UK: Blackwell.

Palmeri, T., Goldinger, S., & Pisoni, D. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19*, 309–328.

Pardo, J. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*, 2382–2393.

Perrachione, T. K., Pierrehumbert, J. B., and Wong, P. C. M. (2009). Differential neural contributions to native- and foreign-language talker identification, *Journal of Experimental Psychology, 35(6)*, 1950–1960.

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

Peschke, Ziegler, Kappes, & Baumgaertner, (2009). A Hierarchical Bayesian Model of Multi-level Phonetic Imitation. *Proceedings of the 27th West Coast Conference on Formal Linguistics* Cascadilla Proceedings Project, Somerville, MA, USA

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175-184.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J.

    Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure*

    (pp. 137-157). Amsterdam, Netherlands: John Benjamins.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology.

    *Language and Speech, 46(*2-3), 115–154.

Pillai, J. J., Araque, J. M, Allison, J. D., Sethuraman, S., Loring, D. W., Thiruvaiyaru, D., Ison,

    C. B., Balan, A., & Lavin, T. (2003). Functional MRI study of semantic and phonological

    language processing in bilingual subjects: preliminary findings. *Neuroimage. 19*(3), 565-

    76.

Pisoni, D. (1993). Long-term memory in speech perception: Some new findings on talker

    variability, speaking rate and perceptual learning. *Speech Communication, 13(*1-2):109–

    125.

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of nonnative speech

    contrasts. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech*

    *perception: The transition from speech sounds to spoken words* (pp. 121-166).

    Cambridge, MA: The MIT Press.

Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. & Fosler-Lussier, E.

    (2007). *Buckeye corpus of conversational speech* (2nd release)

    /www.buckeyecorpus.osu.edu/. Columbus, OH: Department of Psychology, Ohio State

    University.

Posner, M., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review*

    *of Neuroscience, 13*, 25–42.

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication, 41*,

221-231.

Protopapas, A., & Calhoun, B. (2000). Adaptive phonetic training for second language learners. In P. Delcloque (Ed.), *Proceedings of the 2nd International Workshop on Integrating Speech Technology in Language Learning,* pp.31-38. University of Abertay Dundee, U.K

Putman, W. B. and Street, R. L., Jr. (1984). The conception and perception of non-content Speech performance: Implications for speech accommodation theory. *International Journal of the Sociology of Language* , *46*, 97–114.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly, 46*, 842–854.

Schacter, D. L., and Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices, *Journal of Experimental Psychology, 18(5*), 915–930.

Schiller, N.O. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language*, *39*, 484–507.

Schmidt, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics, 11*, 129-158.

Schmidt, A. M. (1996). reported effects of vowel context on cross-language perceptual similarity judgments of Korean and American English consonants by Korean L2 learners of English. *Journal of the Acoustical Society of America, 99,* 3201–3221.

Schmidt, R. W. (2001). Attention. In P. Robinson (Eds.), *Cognition and second language instruction* (pp. 3-32). New York: Cambridge University Press.

Shih, C. (1995) Study of Vowel Variations for a Mandarin Speech Synthesizer. *EUROSPEECH '95*, Madrid, 1807-1810.

Shih, C., Lu, H-Y. D., Sun, L., Huang, J-T., & Packard, J. (2010). An adaptive training program for tone acquisition. Paper presented at *Speech Prosody, 2010,* Chicago, IL.

Showalter, C. E., & Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research, 29*(2), 185-200.

Shriberg, L. & Kent, R.D. (2003). *Clinical phonetics*, 3rd edition, Pearson Education, Boston, Mass, USA,

Silva, D.J. (1992), A Phonetically Based Analysis of /voice/ and /fortis/ in Korean. *Japanese/Korean Linguistics 2*, 164-174

Simon, E., Chambless, D., & Alves, U. K. (2010). Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Language Sciences, 32*, 380–94.

Snoeren, N. D., Gaskell, M. G., & Betta, A. M. D. (2009). The perception of assimilation in newly learned novel words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 542–549.

So, C. (2005). The influence of L1 prosodic background on the learning of Mandarin tones: Patterns of tonal confusion by Cantonese and Japanese naïve listeners. In C. Gurski (Ed.), *Proceedings of the 2005 Canadian Linguistics Association Annual Conference*. Retrieved from http://ling.uwo.ca/publications/CLA-ACL/CLA-ACL2005.htm

Sole, M.-J. (1992). Phonetic and phonological processes: The case of nasalization. *Language and Speech, 35*, 29–43.

Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1992). The effects of speaking rate and amplitude variability on perceptual identification. *The Journal of the Acoustical Society of America, 91(4)*, 2340–2340.

Sommers M. S., Nygaard L. C., and Pisoni D. B. (1994). Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude, *The Journal of the Acoustical Society of America 96(3)*, 1314–1324.

M Sonu, M. Kato, H. Tajima, K. Akahane-Yamada, R. Sagisaka, Y. (2013) Non-native perception and learning of the phonemic length contrast in spoken Japanese: training Korean listeners using words with geminate and singleton phonemes. *Journal of East Asian Linguistics 22 (4)*, 373-398.

Strange, W. (2011) Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics, 39 (4),* 456-466.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners*. Journal of the Acoustical Society of America, 109*, 1691-1704.

Strange, W., & Bohn, O.-S. (1998). Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies. *Journal of the Acoustical Society of America, 104*, 488-504.

Strange W. (1987). Information for vowels in formant transitions. *Journal of the Acoustical Society of America, 26*, 550–557.

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

Strange W., Jenkins J. J., & Johnson T. L. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America, 74*, 695–705.

Summers, W. V. (1988). F1 structure provides information for final-consonant voicing. *Journal of the Acoustical Society of America, 84*, 485–492.

Thomas, R. & Campagna, P. (2010). The Effect of Task Type on Native Speaker Judgments of L2 Accented Speech, *Canadian Acoustics 38 (3*), 122-123.

Thomson, RI, Nearey, T.M. & Derwing, T.M. (2009). A modified statistical pattern recognition approach to measuring the crosslinguistic similarity of Mandarin and English vowels. *The Journal of the Acoustical Society of America 126 (3*), 1447-1460

Trofimovich, P. (2008). What Do Second Language Listeners Know About Spoken Words? Effects of Experience and Attention in Spoken Word Processing. *Journal of Psycholinguist Research* 37:309–329

Truscott, J., & Sharwood Smith, M. (2011). Input, intake, and consciousness: The question for a theoretical foundation. *Studies in Second Language Acquisition, 33,* 497–528

Tse, K. P. (1992). Production and perception of syllable final /n/ and /N/ in Mandarin Chinese: An experimental study. *Studies in English Literature & Linguistics, 18*, 143–156.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics 33*, 263–290.

Van Patten, B. (1996). Input processing and grammar instruction: Theory and research. Processing instruction: An update. *Language Learning, 52,* 755-803.

Van Patten, B. (2002). Processing the content of input-processing and processing instruction research: A response to DeKeyser, Salaberry, Robinson, and Harrington. *Language Learning, 52*, 825-831.

Vaux, Bert and Scott Golder, (2003). *The Harvard Dialect Survey.* Cambridge, MA: Harvard University Linguistics Department.

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of Acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica, 64*(2-3), 122-44.

Wan, I-P., & Jaeger, J. (2003). The phonological representation of Taiwan Mandarin vowels: A psycholinguistic study. *Journal of East Asian Linguistics, 12*(3), 205-257.

Wang, X. (2008). *Perceptual Training for Learning English Vowels: Perception, Production, and Long-term Retention*. Saarbrucken, Germany: VDM Verlag Dr. Muller

Wang, H. & Van Heuven, V. J. (2006). Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. In J. M. Van de Weijer and B. Los (eds.) *Linguistics in the Netherlands 2006*. Amsterdam/Philadelphia: John Benjamins, 237–248.

Wang, M., Koda, K., & Perfetti, C. A. (2003). Alphabetic and nonalphabetic L1 effects in English word identification: A comparison of Korean and Chinese English L2 learners. *Cognition*, *87*(2), 129-149.

Wang, X. (2008). Training Mandarin and Cantonese Speakers to Identify English Vowel Contrasts: Long-Term Retention and Effects on Production. Unpublished doctoral dissertation, Simon Fraser University, BC, Canada.

Wang, X., & Munro, M. J. (2004). Computer-Based Training for Learning English Vowel Contrasts. *System: An International Journal of Educational Technology and Applied Linguistics*, 32(4), 539–552.

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

Wanrooij, K. & Boersma, P. (2013), Distributional training of speech sounds can be done with continuous distributions. *The Journal of the Acoustical Society of America, 133(5)*, DOI:http://dx.doi.org/10.1121/1.4798618

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

White, Y., & Blumstein, M. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. *Journal of Memory and Language, 68*, 362–378.

Wu, Y. (1994). *Mandarin Segmental Phonology*, PhD dissertation, University of Toronto.

Yang, B. (1992). An acoustical study of Korean monophthongs produced by male and female speakers. *Journal of the Acoustical Society of America, 91*, 2280–2283.

Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*, 24, 245–261.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics, 24*, 1−27.

Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics '08*, Paris, France.