Understanding and Addressing Perceptual Challenges for Adult Vietnamese-Speaking ESL Students

by

Anna Cecile Lauzon, B.A.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

August 2009

Understanding and Addressing Perceptual Challenges for Adult Vietnamese-Speaking ESL Students

Approved by
Supervising Committee:

Diane Schallert

L. Alison McGregor

Acknowledgements

This report would not have been possible with out the encouragement and support of many people. My co-supervisors L. Alison McGregor and Diane Schallert were instrumental in helping me achieve this goal. Alison pushed me to explore the many questions that teaching had raised and showed me that teaching pronunciation is both valuable and fascinating. Throughout my studies at the University of Texas, she provided me with opportunities to grow as a scholar and a person. In so many ways, she helped define my academic experience at UT. In addition, Diane offered many useful suggestions and helped me express myself more clearly.

To my friends, thanks for making the experience unforgettable and often humorous. I appreciate everyone who expressed genuine interest in my report topic, especially Eileen Fitzpatrick.

I would also like to thank my family for helping me realize my ambition to go to graduate school and for always believing in me. I could not have done this without you.

August 2009

Abstract

Understanding and Addressing Perceptual Challenges for Adult Vietnamese-Speaking ESL Students

Anna Cecile Lauzon, MA

The University of Texas at Austin, 2009

Supervisors: Diane Schallert, L. Alison McGregor

Pronunciation instruction in the English as a Second Language (ESL) classroom often focuses exclusively on production and fails to address students' perception problems. This report attempts to explain why L2 learners struggle to perceive and produce segments accurately in their L2 and what can be done to help L2 learners overcome these problems. Accordingly, the report explores how L1 experience and segmental differences between Vietnamese and English contribute to these English language learners' perception and production problems. The report also considers instructional methods that can be used to help ESL learners overcome their perception and production challenges and recommends several approaches for addressing segments that are often difficult for Vietnamese learners.

Table of Contents

List of Tables

CHAPTER 1: ADULTS AND SECOND LANGUAGE LEARNING

WE CANNOT UNDERSTAND THEM, BUT WE KEEP PASSING THEM

During my first year teaching English as a Second Language (ESL) in the United States, I taught listening and speaking to a group of adult international students. The majority of the students were Vietnamese, and I noticed that their pronunciation needs were very different from their Russian-, Spanish-, and Mandarin-speaking classmates. I realized this during an activity that was designed to help students with the pronunciation of /f/ and /p/. In practicing minimal pairs, I would say *pan* and *fan* and ask the students to repeat after me. Some of the students seemed comfortable producing the sounds and could repeat back the minimal pair without any problems, but many of the Vietnamese students struggled to produce the sounds accurately. As I tried to help them produce the sounds correctly, several Vietnamese students told me that the pairs sounded the same. I realized that not only did they struggle to produce the sounds differently, but they also could not hear the difference between the sounds. This startling discovery raised the questions that I explore in this report:

Why do adult second language (L2) learners have trouble perceiving sounds and how might it influence their production in their second language?

How can ESL teachers help L2 learners overcome problems perceiving and producing segments in their L2?

I asked my colleagues, twenty-year veteran ESL teachers, if they had had similar problems helping the Vietnamese students improve their listening and speaking skills, and they acknowledged that it was very difficult to address their pronunciation needs in a class of mixed first languages (L1s). One teacher told me that she passed the students

even though she could not understand what they were saying if they completed all of their assignments and worked hard. I was taken aback by this response to the pronunciation needs of what was a large percentage of the population enrolled; even among the experienced ESL teachers, there seemed to be a gap in their understanding of how to deal with the pronunciation needs of native Vietnamese speakers. Yet, I realized that the variety of pronunciation needs of students in a mixed L1 class on top of the course goals (which included improving conversational and listening skills in addition to pronunciation) limited the amount of time that could be spent addressing the Vietnamese-speaking students' pronunciation needs.

This classroom experience made a strong impression on me. It prompted me to want to understand better why the Vietnamese students could not perceive certain sounds, which sounds needed to practiced, and what type of instruction would be most beneficial to them, considering their speech perception problems. Consequently, to address adult Vietnamese students' pronunciation needs better, I knew I would need to understand something about what was at the root of their perception and production problems.

THE RELATIONSHIP BETWEEN L2 PERCEPTION AND PRODUCTION

The relationship between speech perception and production in second language acquisition (SLA) is much debated. Researchers who maintain that speech perception and production are components of one "specialized phonetic module" also believe that improvements in one area will automatically lead to improvements in the other because they are intrinsically related (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997, p. 2299). Consequently, a prevailing idea among L2 researchers and instructors is that L2 learners must first be able to perceive sounds that are in the L2 but not in the learner's L1 accurately, before they can accurately produce those sounds (Sheldon & Strange, 1982).

Other researchers propose that speech perception and production operate independently, and therefore, improvements in one area will not necessarily affect the other (Bradlow et al., 1997).

Sheldon and Strange (1982) investigated native Japanese speakers' perception and production of English /r/ and /l/, which are allophones in Japanese, to determine if perception or discrimination of L2 phones and phonological contrasts was necessary for production of those sounds. Several of the participants in Sheldon and Strange (1982) were able to produce the non-native contrast accurately despite making many errors on perceptual discrimination tests; the authors interpreted these results as evidence that accurate production can precede accurate perception. They emphasized that "perceptual mastery, although often correlated with production accuracy, is not necessarily a prerequisite for, nor a consequence of articulatory mastery" (Sheldon & Strange, 1982, p. 256). Based on their findings, Sheldon and Strange (1982) asserted that when mastery of both perception and production is a course objective, perception and production should be assessed and taught separately to ensure that students improve in both. Although the study provided strong evidence against perception as a requirement for production, it may not be appropriate to generalize its findings about the relationship between perception and production of English /r/-/l/ by native Japanese speakers to other consonant contrasts or vowels, or other L1 speakers.

Bradlow et al. (1997) conducted a similar study that investigated the impact of perceptual training, which consisted of discrimination exercises that used naturally spoken /r/-/l/ minimal word pairs, with feedback on native adult Japanese speakers' production of /r/-/l/ minimal pairs. Most participants showed improvement in their production of the contrast from the pretest to posttest, indicating a "transfer of perceptual learning to aspects of speech production" (Bradlow et al., 1997, p. 2907). Bradlow et al.

(1997) proposed that the transfer of learning of perception to improvements in production indicated that there might be a "unified, common mental representation that underlies both speech perception and speech production" (p. 2308). Yet, the participants' improvements in perception and production proceeded at different rates, which the authors attributed to individual differences that affected the acquisition of the motor commands needed to produce the sounds (Bradlow et al., 1997). Based on their findings, Bradlow et al. (1997) advocated the use of perceptual training to improve L2 learners' pronunciation of nonnative speech sounds.

Although both studies indicated that there is a relationship between perception and production, they offered opposing views on how learning in one area impacted the other. Sheldon and Strange's (1982) findings indicated that perceptual mastery was not a "causative factor in the acquisition of productive skills" (p. 257), whereas Bradlow et al. (1997) indicated that perceptual learning directly transferred to production because perception and production are related. Although unresolved, these disparate views about the relationship between perception and production give ESL instructors much to consider in terms of how perception and production should be approached in pronunciation instruction.

CHAPTER 2: EXPERIENCE-BASED MODELS OF PERCEPTION

THE ROLE OF PERCEPTION IN PHONOLOGICAL ACQUISITION

In order to answer my first question about why my Vietnamese-speaking students could not "hear" certain sounds, I will next review the following four speech perception models: the Native Language Magnet Model, the Perceptual Assimilation Model, the Phonological Interference Model, and the Speech Learning Model. These experience-based models of perception contribute to our understanding of the perception of sounds by providing theoretical explanations for adult L2 speech perception in terms of first language experiences.

Patricia Kuhl's Native Language Magnet Model

Patricia Kuhl's Native Language Magnet (NLM) model theorizes that infants' ambient language experience leads to the formation of native phonetic categories that facilitate native language perception and reduce sensitivity to non-native speech (Kuhl et al., 2008, p. 982). Although infants can initially differentiate phonetic units from any language because of a general auditory processing mechanism, they exhibit language-specific perceptual sensitivities for phonetic units between 6 and 12 months of age and begin to develop phonetic representations based on the input they receive (Iverson et al., 2003, p. B49). With increased experience, prototypes, or the phonetic "representations [that are] most often activated," serve as perceptual magnets, attracting members of the same category and reducing the perception of variation among exemplars of the same category (Kuhl et al., 2008, p. 982). This warping of perception by L1 prototypes is referred to as the *perceptual magnet effect*.

Kuhl et al. (2008) recently revised the NLM model, producing the NLM, expanded (NLM-e), which explains this perceptual warping in terms of biology. According to the NLM-e, infants' exposure to their L1 causes "physical changes in their neural tissue and circuitry," referred to as the native language neural commitment (NLNC) (Kuhl et al., 2008, p. 983). These physical changes in neural networks are based on phonetic patterns of the L1 and ease of L1 processing, but they reduce sensitivity to non-native phonetic input. The NLM-e does not indicate an age at which the "period of optimum sensitivity for phonetic learning" closes, but suggests that infancy is an ideal time for introducing an L2 (Kuhl et al., 2008, p. 994).

While the perceptual magnetic effect is thought to facilitate L1 language acquisition in infants, it may interfere with adult learners' L2 phonetic learning. According to the NLM model, a wide range of sounds (not in an L1) will be perceived as a particular L1 prototype because they will be "magnetically" pulled toward that prototype. This means that normal variations of an L1 sound will be perceived as that L1 sound, but that L2 sounds (to which a person was not exposed during infancy when formation of prototypes occurs) will be perceived as L1 prototypes, especially if the sounds closely resemble that person's native phonemes (Iverson et al., 2003, p. B48). For adult L2 learners, L2 sounds will be distorted and heard as L1 sounds, making it more difficult to learn the language. In addition, Kuhl et al. (2008) proposed that early language exposure also determines which language features are attended to, which means that L2 learners will not attend to features of the L2 that are not relevant to the L1, such as the third formant (F3) for speakers of some L1s, such as Japanese (Iverson et al., 2003, B54).

Catherine Best's Perceptual Assimilation Model

Catherine Best's Perceptual Assimilation Model (PAM) attempts to explain how non-native contrasts are perceived. PAM asserts that listeners assimilate non-native sounds to their native sounds whenever possible based on perceived similarities or differences between the articulatory gestures used to produce the L2 and the L1 sounds (Guion, Flege, Akahane-Yamada, & Pruitt, 2000). In addition, PAM predicts adults' ability to distinguish between non-native contrasts based on how well or how poorly each non-native phoneme assimilates to an L1 phoneme.

If two phonetically similar non-native segments assimilate separately to two different native phonemes, Two Category assimilation (TC) has occurred. In other words the model predicts that the listener will perceive the L2 sounds as two different sounds and be able to tell them apart. When non-native phones assimilate equally well or poorly to a single native phoneme, Single Category (SC) assimilation, PAM predicts that the listener will hear just one sound and have trouble discriminating between the two L2 sounds (Best, McRoberts, & Goodell, 2001). Based on PAM, Best and Strange (1992) predicted that Japanese listeners would assimilate the English /r/-/l/ contrast to a single Japanese category, and this SC assimilation would lead to poor discrimination. In their experiment, the Japanese-speaking participants performed poorly on discrimination tasks with /r/-/l/ contrasts, supporting SC assimilation (Best & Strange, 1992).

According to PAM, when two non-native segments assimilate to a single native phoneme, sometimes one fits better than another, creating a Category Goodness (CG) difference. If this happens, the listener will be better able to discriminate between the non-native sounds better than when they both assimilate equally well to a single category (Best et al., 2001). Best and Strange (1992) found CG difference for the assimilation of English /w/-/r/ contrasts by Japanese listeners. Although both sounds were assimilated

by the Japanese listeners to the category /w/, English /r/ was perceived as a poor exemplar of this native Japanese phoneme. This CG difference between /w/ and /r/ led to better discrimination of the sounds by the Japanese listeners than the /r/-/l/ SC contrast, which did not have a CG difference (Best & Strange, 1992). These findings were consistent with Best et al.'s (2001) simplification of PAM's predictions: "TC>CG>SC" (p. 777). Although PAM predicts that listeners will be most successful at distinguishing non-native contrasts when they assimilate to two different L1 phonemes and least successful when they assimilate to one phoneme, it recognizes that listeners are sensitive to "degrees of similarity and dissimilarity between the non-native and native phones" such as those found in CG differences (Best & Strange, 1992, p. 327).

When segments' articulatory properties are perceived as very different from native phonemes, they are classified as Non-Assimilable (NA) non-speech sounds. In these cases, first language experience will not have an impact on a listener's ability to distinguish between the sounds, and PAM predicts that discrimination of NA should be good if they are perceived to be two different nonspeech sounds (Best et al., 2001). For example, Zulu clicks are easily discriminated by English-speaking listeners because they are perceived as nonspeech sounds and therefore are not associated with native phonemes (Best et al., 2001).

Cynthia Brown's Phonological Interference Model

Cynthia Brown's Phonological Interference Model (PIM) is based on the theory of Feature Geometry. This model proposes that phonemes consist of distinctive features organized into a systemic hierarchy, and this unique structural representation, the phoneme's feature geometry, distinguishes that phoneme from others (Brown, 2000). A phoneme's feature geometry includes any feature that distinguishes it from other L1

phonemes, such as place of articulation. According to Brown (2000), the development of these structural representations during infancy will influence language learning later in life.

As a child is exposed to segments used contrastively, the phonological structures that differentiate between the segments are added to the child's grammar, which is derived from Universal Grammar (UG) (Brown, 2000). UG is believed to be the innate language faculty in the human mind that consists of principles and parameters of language (Cook, 2001). These phonological representations impose boundaries on the child's perceptual system, and as a result, an infant's ability to perceive non-native contrasts decreases as the child's ability to discriminate segments in the L1 increases (Brown, 2000). The phonological structure of the child's native grammar, the principles and parameters of the L1 (including phonological features), becomes the child's filter between acoustic signals and linguistic processing.

Accordingly, adult listeners' feature geometry, based on their L1, determines which non-native contrasts they are able to distinguish and accurately perceive. When an adult listener receives an acoustic signal in the L2, it is broken down into phonetic categories. Then, it is passed through the listener's feature geometry, which further categorizes the signal into one of the listener's phonemic categories. If the listener shares distinguishing features of the non-native contrast with the speaker, the members of the non-native contrast will initially be mapped onto two distinct existing L1 phonemic categories, and the learner will be able to discriminate accurately the non-native contrasts (Brown, 1998). Eventually, the learner will perceive a mismatch between his L1 and L2 phonemes, and with increased exposure, new phonological categories will be established (Brown, 2000). This explains why experienced learners are better at perceiving some contrasts than others, and it highlights the importance of L2 input.

If the listener does not have the same phonemic categories as the speaker, the non-native contrast will be funneled into an existing L1 phonemic category, and the listener will hear the non-native contrast as a single L1 sound that is articulated in a similar way (Brown, 2000). According to PIM, in these cases, the listener will never be able to hear the L2 segments as two distinct phonemes because the listener lacks the underlying feature that distinguishes the sounds (Brown, 1998). Brown (1998) indicated that Japanese learners cannot perceive the /l/-/r/ contrast in English for this reason.

Brown (2000) emphasized that accurate perception of phonemic contrasts is necessary for successful acquisition and that learners will only acquire non-native phonemic contrasts that they perceive as distinct sounds. Having the *features* in native grammar not the *phonemes* is important for accurate perception. In this way, native grammar can limit which non-native contrasts can be successfully acquired (Brown, 2000).

James Flege's Speech Learning Model

James Flege's Speech Learning Model is focused on ultimate attainment in L2 pronunciation and emphasizes that many of L2 learners' production problems are based on problems perceiving L2 sounds (Flege, 1995). The SLM proposes that learners perceptually relate sounds in the L2 to sounds in their L1. The "greater the perceived phonetic dissimilarity between L1 and L2 sounds, the more likely that phonetic differences in the sounds will be discerned" (Flege, 1995, p. 239). When a listener perceives these phonetic differences, new phonetic categories for the L2 sounds may be established. The SLM suggests that eventually production of an L2 sound will correspond to the properties represented by the new phonetic categories (Flege, 1995). If those new categories match a native speaker's L1 sound category then the L2 sounds will

be produced accurately. Flege (1995) emphasized that it is possible that an L2 learner's new categories could be based on different features from those of a native speaker. In those cases, the L2 learner will not produce the sound exactly as a native speaker, but will produce a similar sound based on the features that the learner perceived to be relevant. On the other hand, if the listener cannot perceive differences between the L1 and L2 sounds or two L2 sounds, new categories for the L2 sounds will not form, and a single L1 phonetic category will be used to process the sounds (Flege, 1995). When this happens, the L1 sound and the L2 sound perceived as an L1 sound, will resemble each other in production. For Japanese-speaking English language learners, this might mean that their productions of /r/ and /l/ would be identical because the learner could not hear the difference between them.

In addition to perceived phonetic difference, age of exposure to the L2 can also influence the development of L2 phonetic categories. The SLM asserts that younger learners are more likely to establish new categories than older learners and that the likelihood of an L2 learner perceiving phonetic differences between L1 and L2 sounds decreases as the learner's age of learning increases (Flege, 1995). The SLM does not specify an age at which perception of phonetic differences decreases, but Flege (1995) implied that L2 learners benefit from language exposure prior to adolescence.

COMPARISON OF MODELS

Each of the four experience-based models reviewed offers a unique explanation of L2 learners' perception problems, and each focuses on different aspects of speech as the basis of perception (see Table 1). The NLM indicates that L1 phonemes lead to the development of magnetic prototypes that pull nonnative phonemes toward them, distorting perception, whereas PAM focuses more on the relationship between

articulation and the discrimination of nonnative contrasts, not single phonemes or L1 patterns. The PIM also examines nonnative contrasts but uses feature geometry of the L1 as the basis for determining which contrasts listener will be able to perceive and acquire. The SLM explains L2 perception and production in terms of phonetic distance between the L1 and L2. The variation among these models can be attributed not only to differences in their interpretations of perceptual problems, but also to differences in each model's purposes. For example, SLM is the only model that directly addresses the impact of perception on production because the model was designed to explain L2 pronunciation problems. The other three models were designed to explain how L1 perception influences language growth in children (NLM), cross-language perception (PAM), and acquisition of L2 segments (PIM), and therefore focus almost exclusively on perception.

Despite these differences, all four models suggest that exposure to the L2 is vital for accurate perception of L2 sounds and may have a positive effect on adult L2 perception. Although NLM focuses primarily on infants' language development, Kuhl et al. (2008) acknowledged that successful perceptual training uses exaggerated input that resembles infant-directed speech or motherese. Although the formation of L2 prototypes is beyond the scope of Kuhl's model, it may be possible that sufficient and appropriate input could lead to the development of L2 prototypes as it does for L1 prototypes in infants.

Based on research conducted with experienced L2 learners, Best and Strange (1992) and Flege (1995) also indicated the importance of exposure to L2 sounds, but neither PAM nor the SLM directly addresses how or if training could alter the associations between L1 and L2 sounds. Best and Strange (1992) suggested that "language-specific attunement of phonetic perception may remain somewhat malleable

even in adulthood" (p. 328), but did not indicate whether or how established perceptual assimilation patterns can change once L2 sounds have been assimilated to L1 sounds (Escudero, 2007), which often occurs with L2 learners. Despite PAM's emphasis on the role of the "listener's knowledge (whether implicit or explicit)" of articulatory gestures in determining phonetic distance (Best et al., 2001, p. 777), the model does not address whether explicit instruction in place and manner of articulation could have a positive impact on perception (Guion et al., 2000). In addition, Best and Strange (1992) suggested that more research is needed to understand what factors can lead to adjustments in perceptual abilities, and they identified perceptual training that contrasts non-native segment with the nearest native phoneme as a method that merits investigation.

Despite the SLM's assertions about age of learning, Flege (1995) acknowledged that several studies have indicated that experience can alter the learner's phonetic categories, and consequently, impact learners' abilities to perceive and produce L2 sounds. Flege (1995) indicated that with experience, learners may gradually begin to recognize differences between L2 sounds that they initially perceived as L1 sounds, which may in turn, lead to the development of new phonetic category representations for the L2 sounds. The SLM indicates that new phonetic categories form only when learners discern phonetic differences between the L1 and L2 sounds, but Flege (1995) did not explain how L2 learners gauge phonetic distance or how learners could be made aware of this distance (Escudero, 2007). Rather, Flege (1995) suggested that more studies of longitudinal changes in perception and production are needed.

Like PAM, PIM can be used to predict which contrasts L2 learners will struggle with, but PIM, in contrast, suggests that given enough input some learners, provided they have the right features in their L1, can be made aware of differences between certain L1

and L2 contrasts and can acquire the L2 phonemes. Yet, PIM is extremely pessimistic about some L2 learners' ability to learn certain non-native contrasts. According to Brown (1998), learners who do not have the features associated with an L2 phoneme cannot accurately perceive that sound and in turn cannot acquire it. The model does not address production or whether learners could be made aware of features that are not in their L1, such as place of articulation, and taught to produce L2 sounds using those places of articulation despite not being able to perceive those sounds accurately. Although all four models indicate that the L1 can interfere with the accurate perception and even learning of some L2 sounds, PIM is the only model that asserts that certain first languages will entirely prevent the acquisition of specific L2 sounds (when the L1 and L2 do not share phonemic categories).

Table 1

Comparison of Experience-Based Models

	NLM	**PAM**	**PIM**	**SLM**
How does the L1 influence L2 perception?	L1 prototypes act as perceptual magnets. Native language neural commitment reduces sensitivity to non-native input near L1 prototypes (Kuhl et al., 2008)	L2 perception based on detection of articulatory-phonetic similarities with L1. Accuracy in discrimination related to how L2 sounds of non-native contrasts assimilate to L1 sounds (Best et al., 2001).	L2 input perceived in terms of L1 phonological categories in native grammar. When lacking feature in L1, leads to distortion of input (Brown, 1998). Capacity to perceive all non-native contrasts change as child develops (Brown, 2000).	L2 sounds can be perceived as L1 sounds because of failure to discern phonetic differences between the sounds when L1 phonological system filters out properties of L2 sounds (Flege, 1995).
Why was the model developed?	To explain how native phonetic perception predicts language growth in children (Kuhl et al., 2008).	To predict cross-language perception, specifically, assimilation and discrimination differences for non-native contrasts (Best et al., 2001).	To explain influence of L1 phonology on acquisition of L2 segments (Escudero, 2007).	To explain age-related limits on the production of L2 sounds and ultimate attainment in L2 pronunciation (Flege, 1995).
Basis for perceptual differences?	Distributional frequencies & patterns of acoustic signal and exaggerated acoustic cues (Kuhl et al., 2008).	Phonetic-articulatory (Best et al., 2001).	Phonological structure of feature geometry, including place of articulation (Brown, 2000).	Phonetic features (Flege, 1995).
Reaction to the Critical Period (CP) Hypothesis, age of learning, and/or Universal Grammar (UG)?	Neural commitment as the cause of CP phenomenon. CP closes when phonetic categories reach stability (Kuhl et al., 2008). Decline in non-native speech perception occurs between 6 and 12 months (Kuhl et al., 2008). Humans use strategies and recognize patterns, not UG (Kuhl, 2000).	Language-specific attunement of phonetic perception may remain malleable for adults (Best & Strange, 1992). Experience may help adults categorize and discriminate using phonemic and phonetic properties of L2 (Best & Strange, 1992).	Does not directly address CPH, but allows that exposure to L1 reduces sensitivity to non-native contrasts during childhood (Brown, 1998). Perception problems, not lack of access to UG, lead to adult learners' problems acquiring L2 phonemes (Brown, 1998).	Mechanisms and processes used to learn L1 sounds remain intact and can be used in L2 learning (Flege, 1995). Older learners have developed L1 categories but can form new categories if perceive phonetic differences between L1 and L2 sounds (Flege, 1995). Learners under age 10 better able to observe L2 phonetic detail and gestures (Flege, 1995).

Table 1 (continued)

	NLM	PAM	PIM	SLM
Limitations	Does not explain how adult learner can create new mappings for the L2 (Escudero, 2007).	Does not explain how to change established perceptual assimilation patterns of L2 (Escudero, 2007).	Does not account for how learners could learn features not in native grammar (Escudero, 2007).	Does not explain how distance between phonetic features can be measured (Escudero, 2007).
Why do students struggle with perception of L2 sounds or contrasts?	If sound near L1 prototype, that prototype pulls L2 sounds to L1 prototype and perceived as L1 category.	Contrast may be assimilated onto a single category and perceived as equally good fit.	When L1 and L2 do not share features, contrast is perceived in terms of L1 and is mapped onto L1 categories.	If both sounds identified in terms of single L1 category, learner has trouble perceiving as different.
Can novel sounds or contrasts be perceived?	Yes, can be trained to perceive L2 contrasts by using exaggerated input and exposure to different speakers (Iverson et al., 2003). Not automatically form new neural structure as children do (Kuhl et al., 2008).	Yes, but depends on how contrasts are assimilated. Two Category assimilation best for discrimination. Non-speech sounds can also be perceived. (Best et al., 2001).	Yes, but only if learner has same feature in L1. Otherwise, accurate perception of sounds will be blocked (Brown, 1998).	Yes, if learner perceives that L2 sound differs phonetically from closest L1 sound (Flege, 1995).
Can novel sounds be acquired or produced?	Perception problems can lead to accented adult speech because it is hard to unlearn L1 motor speech (Kuhl, 2000). If speaker does not have L2 prototypes, cannot produce L2 without accent (Iverson, et al., 2003).	Does not address production.	A learner can only acquire L2 sounds if can accurately perceive them and can only perceive them if have same feature geometry (Brown, 1998). If the learner cannot perceive the sound, the learner will produce it using most similarly articulated L1 sound (Brown, 2000).	Accuracy of production based on how sounds perceived. Can form new categories if perceived as very different from L1 sound. Production is based on properties in category formed (Flege, 1995).
Pedagogical Implications	Best results when exposure to L2 sounds happens in the first year of life (Kuhl et al., 2008). Teach adults contrasts with exaggerated phonetic cues and with many different speakers (Iverson, 2003).	Model predicts perception but does not indicate how to change it. Suggests exposure to L2 input may be beneficial, but unclear how it would impact perceptual assimilation patterns (Best & Strange, 1992).	Exposure to L2 sounds benefits learners who have the same features in their L1s (as L2 sounds) and can form new phonological categories, but not learners without the features (Brown, 2000).	Teach L2 before age 10 for best results. Exposure to L2 can lead to perception of phonetic differences, development of new L2 categories, and more accurate production (Flege, 1995).

CHAPTER 3: PHONOLOGICAL COMPARISON OF ENGLISH AND VIETNAMESE SEGMENTS

The experience-based models of perception suggest that the learners' L1 experiences have a negative effect on their ability to perceive and, in the case of the SLM, produce English segments that are not part of their L1. Phonologically, English and Vietnamese share some segments, yet each language has segments that are not found in the other. The following comparison of Vietnamese and English identifies segmental differences between the languages that may lead to perception problems and pronunciation errors for adult native Vietnamese-speaking English language learners. These differences and similarities not only affect how learners perceive the segments, but also "how native Vietnamese speakers produce English sounds and how the productions are perceived by English speakers" (Hwa-Froelich, Hodson, & Edwards, 2002, p. 266).

CONSONANTS

Although Vietnamese and English share many consonants, there are fewer final consonants in Vietnamese, and those that are shared by both languages are often produced differently or are found in different positions within syllables (Hwa-Froelich et al., 2002). For example, final consonants in Vietnamese are limited to either a voiceless stop /p/, /t/, or /k/ or a nasal /m/, /n/, or /ŋ/ (Hwa-Froelich et al., 2002). In Vietnamese, /p/, /t/, and /k/ are unreleased, and Vietnamese English language learners may substitute their unreleased /p/, /t/, and /k/ for English's similar released consonants in words ending in these consonants, such as *stop*, *boat*, and *pack* (Hwa-Froelich et al., 2002).

Because Vietnamese does not have voiced stops at the ends of syllables, Vietnamese speakers' productions of voiced stops found in English may strongly

resemble the unreleased voiceless stops found in Vietnamese (Avery & Ehrlich, 1992). This means that Vietnamese speakers' productions of English voiced stops /b/, /d/, and /g/ are often indistinguishable from the voiceless stop /p/, /t/, and /k/, and in some cases, they cannot even be perceived by a native English speaker (due to the voiceless stops being unreleased). For example, a Vietnamese speakers' production of *made* may sound like *mate* or *may* to a native English speaker. Other consonants that do not appear in the word-final position in Vietnamese are often dropped, including the fricatives /f/, /v/, /θ/, /ð/, /s/, /z/, /ʃ/, and /ʒ/ (Avery & Ehrlich, 1992).

In addition to final consonants, initial consonants can also pose problems for Vietnamese-speaking learners of English. These non-native English speakers often omit or substitute word-initial consonants that are not a part of Vietnamese with Vietnamese segments. For example, /t/ and /k/ in the initial position are unaspirated in Vietnamese, whereas in English they are never unaspirated at the beginning of a syllable (Hwa-Froelich et al., 2002). Therefore, Vietnamese speakers' productions of /t/ and /k/ are often perceived as English /d/ and /f/ (Hwa-Froelich et al., 2002). This also occurs with the initial interdental fricatives /θ/ and /ð/, where /θ/ is often replaced by /t/, and /ð/ is replaced by /d/ (Avery & Ehrlich, 1992).

Consonant dropping and substitution of English segments with Vietnamese segments can create serious intelligibility problems. For example, when I was teaching a listening and speaking unit on money, a Vietnamese student responded to a question about loans. She said something about *cash*, but I heard /*ka*/. I thought she might be talking about a car loan. No. She repeated herself several times until another student told me what she was saying. This type of misunderstanding can be embarrassing and frustrating for L2 learners and may discourage them from attempting to communicate in the L2 in the future.

Consonant clusters

One major difference between English and Vietnamese is that English syllables can include consonant clusters, whereas Vietnamese only includes the consonant cluster /kw/ and the consonant cluster /h/ plus a consonant (Hwa-Froelich et al., 2002). Consonant clusters are more than one consonant together at the beginning (initial cluster) or end (final cluster) of a syllable (Avery & Ehrlich, 1992). Consonant clusters can be difficult for English language learners, and native Vietnamese speakers often delete one or more of the consonants in a cluster (Avery & Ehrlich, 1992).

Vowels

Generally, Vietnamese speakers do not have too much difficulty producing English vowels. English and Vietnamese have seven vowels in common. Vietnamese uses five single vowels and seven triphthongs that are not part of English, and English contains three single vowel sounds, /I/, /ə/, and /ʌ/, that are not in Vietnamese (Tang, 2007). Yet, despite the many vowel distinctions in Vietnamese, English tense/lax vowel pairs can be difficult for Vietnamese learners (Avery & Ehrlich, 1992). For example, distinguishing between the tense/lax pairs /i/-/I/ and /u/-/ʊ/ can be problematic for Vietnamese speakers (Hwa-Froelich et al., 2002).

The features identified above can pose difficulties for Vietnamese-speaking English language learners and can significantly hinder comprehensibility and intelligibility. The substitution errors and consonant dropping observed by teachers of native Vietnamese-speaking English language learners may be attributed to the imperfect assimilation of L2 sounds to L1 sound categories, as described by the SLM (Flege, 1995). If these errors are interpreted from the perspective of Flege's model, the Vietnamese learners are producing the sounds in terms of how they are perceived, which in many instances is as their L1 sounds. For teachers of Vietnamese-speaking English language

learners, understanding the source of the learners' errors may help them better address their perception and production problems.

Although this report focuses specifically on segmental differences between the two languages, an analysis of differences between Vietnamese and English at the suprasegmental level would provide additional insight into other pronunciation challenges for Vietnamese-speaking English language learners. It would also illuminate additional features of pronunciation that merit instruction but are beyond the scope of this report.

CHAPTER 4: IMPLICATIONS FOR INSTRUCTION

INSTRUCTIONAL METHODS FOR L2 LEARNERS WITH PERCEPTION PROBLEMS

To answer my second question about how to help L2 learners overcome problems perceiving and producing segments in their L2, I will explore the implications of the experience-based models of perception on instruction. I will also include instructional methods such as the use of acoustic input, explicit instruction, visual input, practice, and feedback, all of which, I will argue, play a critical role in raising the adult learners' awareness and will potentially develop their perception and production of target L2 sounds.

Although all of the models maintain that L1 experience hinders adult L2 perception, they also indicate that adults may benefit from L2 input, either exaggerated input (Kuhl et al., 2008) or input that contrasts L1 and L2 sounds (Best & Strange, 1992). In other words, input that makes L2 learners aware of relevant features is essential and useful. In addition, several of the models highlighted the importance of articulation (PAM) and features (PIM) on perception. Although neither of these models indicated whether making L2 learners aware of these features of speech production could improve perception or production, research in this area has had positive results (Hardison, 2003). While learners may not initially be able to perceive L2 input accurately, instructional methods that highlight aspects of speech that contribute to perception can be used to improve L2 perception and production.

Acoustic input

Although none of the models explicitly identifies what type of L2 input could benefit adult L2 learners with perception problems, several studies have indicated that L2

input can be modified to raise the learners' consciousness, increase their rate of learning, and help them attend to the relevant contrast (N. Ellis, 2005). Adult L2 learners with different L1s attend to different dimensions of the same English stimulus, and oftentimes, they attend to different dimensions than those to which native speakers attend (Kuhl, 2000). Iverson et al.'s (2003) findings indicated that Japanese learners of English were able to discern differences between /r/-/l/ contrasts, but noted that they attended to the wrong information in the acoustic signal. Based on these findings, Iverson et al. (2003) recommended exposing adult L2 learners to large stimuli sets because their "variability provides information about which cues are most robust and trains individuals to ignore irrelevant variation" (p. B54).

Stimulus sets that use exaggerated acoustic cues and expose learners to many different speakers and many exemplars of the feature, which are characteristics of the language to which infants are exposed (Kuhl, 2000, p. 11855), have been found to be effective components of perceptual training. These stimulus sets are thought to help adult L2 learners attend to the relevant acoustical cues. Using input with exaggerated acoustic cues may also be a way to overcome the magnetic pull of L1 phonetic categories, as suggested by the NLM model, and unlike other L2 input, it prevents the reinforcement of associations between L1 and L2 phonemes that "dig [L2 learners] ever deeper into the hole begun and subsequently entrenched by their L1" (N. Ellis, 2005, p. 326).

In addition to using exaggerated stimuli to develop L2 learners' perceptual abilities, L2 instructors can provide "the phonemes in natural words" and sequence discrimination tasks so that they are initially easy and become increasingly difficult as students progress (N. Ellis, 2005, p. 328). This means starting with non-native contrasts that are in positions that are easy for the learners to discriminate. Once learners can successfully discriminate these contrasts, the same contrasts could be used but in another

syllable position that is more difficult for the learners to perceive. It could also mean starting with highly exaggerated input and moving to less exaggerated input after learners successfully discriminate the exaggerated non-native contrasts. A study by McCandliss, Fiez, Protopapas, Conway, and McClelland (2002) found that adaptive training that initially used stimuli with exaggerated acoustical cues and increasingly used less exaggerated stimuli as participants successfully responded to the discrimination tasks was more effective than training that used a fixed set of stimuli. With this type of training, students can build on what they have learned without reinforcing their initial perception problems (N. Ellis, 2005).

Although developing stimulus sets like those described by Iverson et al. (2003), Kuhl (2000), N. Ellis (2005), and McCandliss et al. (2002) is probably not feasible for most L2 instructors due to time constraints and limited resources, the studies do have implications for the L2 classroom. For example, it may be possible for L2 teachers to record several different native speakers reading a list of words and exaggerating the sounds being studied. When developing the word lists, the teacher should consider in which position (initial, medial, final, or within clusters) the sound is easiest for the L2 learners to perceive, which will depend on the learners' L1 phonology (Hardison, 2003). The lessons should progress from the sounds and sound positions that are easiest for the learners to more difficult sounds and positions once the material presented in earlier lessons is mastered (N. Ellis, 2005), and the use of exaggeration should be reduced as the lessons progress (McCandliss et al., 2002). This technique could be used in conjunction with the other instructional techniques mentioned in this chapter, including explicit instruction and visual input.

Explicit instruction

The experience-based models of perception imply that aspects of an L2 (among them L2 sounds, articulatory gestures, and features) cannot be learned through implicit mechanisms alone as they are during infancy and childhood because adult learners' L1 experiences impinge on their abilities to perceive sounds accurately and to attend to the relevant features of the L2 input. Because of these perception problems, they do not accurately perceive L2 sounds. This is consistent with the position that language processing happens automatically, to the detriment of SLA (N. Ellis, 2005). According to N. Ellis (2005), unless a learner's attention is drawn to differences between the L1 and L2, L2 input will continue to be processed "in the same old L1 way" (p. 327). Schmidt (2001) hypothesized that *noticing* is necessary for L2 acquisition and that explicit instruction can be used to enhance noticing, particularly for elements of L2 production that are unlikely to be learned implicitly (as cited in DeKeyser, 2003, p. 331).

Explicit instruction, which provides learners with information about the rules underlying the phonological input (Hulstijn, 2005), can be helpful for teaching features that are "too abstract, too distant, too rare, too unreliable, or too hard to notice" (DeKeyser, 2003, p. 334) or are "acquired very slowly […] from implicit processes alone" (N. Ellis, 2005, p. 307). N. Ellis (2005) maintains that explicit instruction can speed SLA, is more effective than implicit instruction, and leads to durable results. His emphasis on the advantages of explicit instruction in teaching features that cannot be learned through implicit mechanisms alone is particularly relevant to pronunciation instruction for adult L2 learners with perception problems.

Michas and Berry (1994) noted the benefits of explicit pronunciation instruction for adult learners. Their study compared the impact of explicit and implicit rule presentation on participant performance on posttests that measured ability to determine

the pronunciation of unfamiliar words and to judge whether a word and its associated pronunciation were accurate. They found that participants who were presented with pronunciation rules, which consisted of the Greek letters and their corresponding pronunciation, and an exemplar, the rule group, performed better on pronunciation posttests than the participants who simply received the exemplar word list and their pronunciations, the natural group. The rule group significantly outperformed the natural group in a test with unfamiliar stimuli, which indicated that they were better able to transfer their knowledge of rules to new words than the group that experienced the "more naturalistic learning condition" (Michas & Berry, 1994, p. 370). Within the rule group, the group who received the rules visually outperformed the group who heard the rules. The results of Michas and Berry (1994) suggested that explicit pronunciation instruction, particularly that which uses visual input, is more effective than implicit pronunciation instruction.

The experience-based models of perception and recent studies (Michas & Berry, 1994; Best & Strange, 1992; Iverson et al., 2003) suggest the need for raising L2 learners' awareness of aspects of the L2 they cannot hear through explicit instruction and providing them with strategies or methods for producing unfamiliar L2 sounds or features (Acton, 1984). If L2 learners are provided with explicit instruction about how sounds are produced, visual input, and opportunities for practice and feedback, they can learn to produce the sounds and possibly overcome their perception problems.

Explicit instruction in place and manner of articulation

PIM and PAM emphasized the importance of articulation and features to perception, which suggests the need for more explicit pronunciation teaching of place and manner of articulation for students who cannot perceive these differences in L1 and L2

production. Although this instruction may not improve perception, which PIM suggests is impossible when the L1 and L2 do not share features, it may improve L2 learners' production of unfamiliar L2 segments that consist of those features.

Explicit pronunciation instruction in place and manner of articulation can include verbal explanations, phonological rules, and visual aids that explain where and how segments are produced and even when they are produced that way. Three-dimensional computer-animated images of faces with transparent skin (Hardison, 2007), "sammy diagrams" of a cross-section of a head that include the position of tongue, teeth, and lips during segment production, and exaggerated teacher-led demonstrations are among the tools used to show students how segments are produced (Avery & Ehrlich, 1992). When students' L1s and L2s do not share features or articulatory characteristics, explanations of how segments are produced may make students more aware of the features and may improve their perception and production.

Although Jones et al. (1994) found that exposure to phonological rules did not always result in more accurate pronunciation, their findings indicated that participants were "better equipped to assess their own speech and more aware of their particular pronunciation problems" (as cited in Jones, 1997, p. 108). The participants' ability to monitor their own production using explicit knowledge indicated that they noticed, and possibly even perceived, their pronunciation problems and could use this information to correct their production errors.

To teach students to produce accurately sounds they struggle to perceive, Acton (1984) recommended taking a kinesthetic and tactile approach, "focus[ing] on how the sound ought to 'feel'—not on how it ought to sound" (p. 78). Acton (1984) stressed that "as long as [the students] can 'remember' the physical sensations, they can be led to practice the 'feel' of the distinction until they, themselves, begin to hear it" (p. 78).

Acton (1984) implied that helping students develop an understanding of L2 segment production may not only improve the production but also the perception of adults L2 learners whose L1 phonology inhibits their ability to notice differences between native speakers' productions and their own.

Visual Input

Adult L2 learners often struggle to *hear* differences between L2 segments and differences between their L1 segments and L2 segments, but that does not mean they cannot *see* differences between sounds. Because speech is a "multimodal phenomenon" as demonstrated by the McGurk effect (an experiment in which speech gestures that were not consistent with the auditory information presented influenced which sound participants reported hearing), foreign language educators should use all modalities to their advantage (Navarra & Soto-Faraco, 2007). For example, L2 teachers should ensure that learners have many opportunities not just to hear sounds being produced but also to see them being produced, see how they are producing them, and contrast their own production with a native speaker's. This can explicitly draw L2 learners' attention to the movements that make the target sounds.

Research has indicated that L2 learners can benefit from audio-visual perceptual training (Hardison, 2007). Hardison (2003) conducted an experiment in which groups of Japanese and Korean participants received either auditory only or combined auditory and visual training. The groups were exposed to minimal pair /r/-/l/ in initial, initial cluster, final, and final cluster positions. During each 30-minute training session, the auditory group listened to a speaker and the auditory-visual group watched a video of a speaker reading a minimal pair list; then, the participants selected which segment they heard. They received immediate feedback on their selections. The students heard five different

speakers read each list a total of three times over the course of three weeks. The results of the posttest indicated "training two modalities (auditory and visual) simultaneously was superior in improving perceptual accuracy to training only one" (Hardison, 2003, p. 517). The visual information, the video of the speaker, contributed most significantly to accurate perception of the segments that were in particularly challenging phonetic environments for the participants based on their L1 phonology (Hardison, 2003). In addition, the participants were able to generalize their training to new words and to new speakers, and their production of the segments improved as well. Hardison (2003) highlighted that these improvements in production happened in "the absence of explicit production instruction," through perceptual training alone (p. 515).

The findings of the experiments conducted by Hardison (2003) have strong implications for pronunciation instruction. The experiments involved perception training, but they led to significant improvements in both perception and production. These findings suggest that students' abilities to perceive and produce segments that they have difficulty perceiving could improve through auditory-visual perceptual training. Hardison (2007) suggested that focused perceptual training that uses multiple exemplars of phonemes or phonemic contrasts in conjunction with feedback and repetition "allow[s] learners to attend to those stimulus features that provide useful input" and "increases the salience and information value of important auditory and visual characteristics" (p. 148).

Based on these findings, L2 teachers should consider using visual input in the classroom, and this training could take many different forms. Teachers can provide their students with videos of speakers producing words so that students can see how the words are produced. They can also ensure that their students in class can see their faces as they produce sounds that the students struggle to perceive and produce. In addition, teachers can provide students with mirrors or make use of Photo Booth (an Apple software) so

that the students can observe themselves as they practice producing unfamiliar sounds, taking advantage of this easily obtained visual feedback.

Practice & Feedback

Although the experience-based models of perception do not address how practice can improve accuracy in L2 production, several theories of second language acquisition do address this issue. Krashen (1985), in what is "referred to as the non-interface hypothesis," advocated large quantities of comprehensible input over the presentation of explicit rules and systemic practice (as cited in DeKeyser, 2003, p. 328), whereas other researchers recognize the benefits of practice. For example, some SLA researchers view practice as a way to "bridge the gap between explicit knowledge and use" or to improve fluency (DeKeyser, 2003, p. 328). Skill-learning theory asserts that when explicit knowledge is "proceduralized through practice," it becomes implicit knowledge, and it is this implicit knowledge that is processed automatically and used during fluent speech (R. Ellis, 2005, p. 214).

For instructors working with learners who struggle to perceive L2 input accurately, the non-interface hypothesis does not offer much guidance. In contrast, research does offer some evidence to support the role of practice in form-focused instruction. N. Ellis (2005) noted that the "balance of experimental findings supports the effectiveness for SLA of encouraging learners to produce output" (p. 337). N. Ellis (2005) emphasized that experiments that involved explicit focus on forms followed by output practice "encouraged the noticing of task-relevant linguistic forms" and showed "evidence of enhanced rule-learning" (p. 337).

Speaking in the L2 also provides learners with opportunities to monitor their own output (N. Ellis, 2005). The learner's monitor "comes into play after the utterance is

initiated and produced by the acquired system and […] can then be used to make changes to the form of the utterance if it seems erroneous" (N. Ellis, 2005, p. 330). The experience-based models of perception would seem to suggest that learners cannot monitor their output if they cannot perceive the sound. Acton (1984), however, suggested that learners can monitor their production kinesthetically. L2 learners can be made aware of "what it feels like to talk and manipulate sound quality" in the L2, and based on this kinesthetic feedback, they can modify their L2 production (Acton, 1984, p. 78). By teaching students how producing L2 segments should feel, L2 instructors give learners a tool and strategy that they can use both inside the classroom and outside the classroom, particularly during a break down in communication.

Research by Michas and Berry (1994) has also identified practice as critical for learning pronunciation rules. Michas and Berry (1994) found that "explicit study of the rules is only beneficial when followed by the experience of applying the rules in the subsequent learning phase" (p. 376). The participants who were given the rules and opportunities to practice them were better able to apply the rules and identify correct and incorrect instances of their application than the group that studied the words and learned the rules to criterion (Michas & Berry, 1994, p. 376). These results highlight the importance of explicit instruction combined with practice.

Michas and Berry (1994) suggested that practice may help with automatization of pronunciation rules, in this case those regarding letter and sound correspondences, because corrective feedback received may lead to attention being focused on the rules presented and subsequently in application of those rules. Other researchers have also found that the use of immediate feedback, rather than delayed feedback, increases the likelihood that the association between cues and their outcomes will be learned (N. Ellis, 2005, p. 328).

McCandliss et al. (2002) also noted the benefits of immediate feedback during pronunciation training with adult Japanese participants. In their experiment, participants in the experimental groups received training that used perceptual discrimination tasks with the minimal pair /r/-/l/. Although the group who received adaptive training with exaggerated acoustical cues showed evidence of transfer to novel stimuli after six days of training, the training that included immediate feedback on perceptual discrimination "produced clear evidence of transfer, even after 3 days" (McCandliss et al., 2002, p. 105). McCandliss et al. (2002) asserted that training with feedback "appears to produce evidence of categorical perception" and suggested that such positive results after just three 20-minute training sessions with a computer "underscores the point that adult language learners maintain considerable plasticity in their ability to learn perceptual speech contrasts" (p. 106). The study suggested that using highly exaggerated contrasts and feedback, which are not typical of most communicative L2 contexts, as well as focusing on a single contrasting stimulus pair, can lead to rapid learning.

INSTRUCTIONAL RECOMMENDATIONS FOR VIETNAMESE-SPEAKING ENGLISH LANGUAGE LEARNERS

Because Vietnamese and English do not have all the same segments and because the segments that are shared are found in different positions within syllables (Hwa-Froelich et al., 2002), native Vietnamese-speaking English language learners may not perceive these sounds. In turn, they may omit, reduce, or replace these segments with similar Vietnamese ones, compromising their intelligibility and comprehensibility. Based on the experience-based models of perception and studies of effective forms of perceptual training and instruction for L2 learners with perception problems, I recommend that instructors of native Vietnamese-speaking English language learners

teach the items listed below using enhanced acoustic input, explicit instruction, and visual input combined with ample opportunities for feedback and practice. I believe teaching learners how these segments are produced and helping them attend to appropriate visual and acoustical cues through perceptual training will lead to improvements both in perception and production.

Specific segment groups and ideas concerning how to raise learners' awareness of them are listed below:

Aspiration: Because the voiceless stops /p/, /t/, and /k/ are not aspirated in Vietnamese, native Vietnamese-speaking English language learners may need instructions in how to produce these segments in syllable initial and final positions. I recommend that teachers explain that *air* is released with the consonant and that they demonstrate this by holding a piece of paper in front of their mouths (Avery & Ehrlich, 1992). The paper should move as the air is released. Students can practice this as well and kinesthetically and visually monitor their own production. Teachers can also contrast a student's production of the similar unaspirated Vietnamese segments with the teachers' productions of English /p/, /t/, and /k/ and ask students to identify visual and acoustic differences.

Voiced and voiceless stops: The voiced stops /b/, /d/, and /g/ and voiceless stops /p/, /t/, and /k/ are among the final consonants that are difficult for Vietnamese speakers (Avery & Ehrlich, 1992). Omission of any final consonant can hinder intelligibility and can significantly change a speaker's meaning. For example, omission of the final stops /t/ and /d/, which are grammatical morphemes that indicate past tense can alter the meaning of what is said. Explicit instruction can be used to raise learners' awareness of differences between the production of voiced and voiceless stops; the teacher can have students put one hand on their throats to determine if they are voicing the sound or not.

When they produce voiced stops, students should feel their vocal cords vibrating as they touch their throats. They can use this method to kinesthetically monitor their production of these stops as they practice them. Audio-visual input that includes minimal pairs, such as those with /b/-/p/, that contrast voiced and voiceless stops could also be used to improve perception and production.

Fricatives: While not often a problem in the syllable initial position, syllable final fricatives can be difficult for many Vietnamese speakers (Avery & Ehrlich, 1992). If students can produce fricatives in the initial position, a teacher can explain that they are the same sounds as those in the final position. The teacher can create linking exercises with fricatives appearing at the end of words that occur before words beginning with vowels; these exercises give students practice saying final fricatives in an environment that resembles a word-initial position (Avery & Ehrlich, 1992). Practice producing words with the grammatical endings /s/ and /z/, found in the plural, possessive, and third-person singular, would also be beneficial for students once they understand how to produce the sounds (Avery & Ehrlich, 1992). For students who struggle to produce fricatives, instruction in place and manner of articulation will be helpful.

Consonant clusters: To help students with consonant clusters, a teacher can simplify the cluster to a single consonant, and then add consonants to it as students become comfortable with the increasingly complex words, for example moving from *brad* to *brand* to *brands* (Avery & Ehrlich, 1992). Audio-visual discrimination tasks and perceptual training could also be used with the same lists of words. Students should also be made aware of the consonant clusters that native English speakers often simplify in connected speech in order to address the intelligibility problems this may create for non-native listeners (Avery & Ehrlich, 1992).

Tense/lax vowels: Although Vietnamese and English share many vowels, the tense/lax distinction can be challenging for some learners. Exaggerated input and modeling of tense/lax pairs may provide learners with helpful information regarding differences in production and length. To further emphasize length, teachers can pull a rubber band as they say tense vowels. By making learners aware of the semi-vowels found at the end of tense vowels, instructors may help learners notice this distinctive feature (Avery & Ehrlich, 1992). For example, instructors can explain that /e/ in *rain* is not a single sound; rather, it consists of /e/ and the semi-vowel /j/, and is often transcribed as /ej/ or /eI/. To make this sound, the tongue must move from the mid front vowel position to the high front vowel position. In addition, instructors can focus on the physical differences during production of tense and lax vowels. For example, the students could be encouraged to touch their neck to feel if they are tensing or relaxing these muscles or to focus on whether the muscles in their face feel tense or relaxed.

These instructional methods should be combined with opportunities to practice the sounds in words, and students should receive feedback on their abilities to perceive and produce the sounds. Although feedback on production may include comments on student's intelligibility and suggestions for how to alter articulation, feedback can also be given on perception through evaluation of discrimination tasks.

Although the importance of perception is often overlooked in many ESL classrooms, perception contributes significantly to students' listening and speaking skills. By better understanding why students cannot hear sounds, teachers can better address a critical factor in their students' oral skill development needs. Using cursory knowledge of their students' L1s and a wide range of teaching methods and materials, including audio-visual input, exaggerated input, explicit instruction in place and manner of

articulation, feedback, and practice, teachers can help their students develop their perceptual and productive abilities in the L2 over time.

References

Acton, W. (1984). Changing fossilized pronunciation. *TESOL Quarterly, 18*(1), 71-85.

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

Best, C. T., McRoberts, G. W., & 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*(2), 775-794.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*(3), 305-31.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & 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), 2299-2310.

Brown, C. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research, 14*(2), 136-193.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In John Archibald (Ed.), *Second language acquisition & linguistic theory* (pp. 4-63). Oxford: Blackwell.

Cook, Vivian. (2001). *Second language learning and language teaching*. New York: Oxford University Press.

DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 313-348). Malden, MA: Blackwell.

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*(2), 305-352.

Ellis, R. (2005). Principles of instructed language learning. *System, 33(2),* 209-224.

Escudero, P. (2007). Second language phonology: The role of perception. In M. Pennington (Ed.), *Phonology in context* (pp. 109-134). New York: Palgrave Macmillan.

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

Gorsuch, G. J. (2001). Testing textbook theories and tests: The case of suprasegmentals in a pronunciation textbook. *System, 29*(1), 119-136.

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

Hardison, D. M. (1999). Bimodal speech perception by native and nonnative speakers of English: Factors influencing the McGurk Effect. *Language Learning*, *49*, (Suppl. 1), 212-283.

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

Hardison, D. M. (2007). The visual element in phonological perception. In M. C. Pennington (Ed.), *Phonology in Context* (pp. 135-158). Basingstoke, Great Britain: Palgrave Macmillan.

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in Second Language Acquisition, 27*(2), 129-140.

Hwa-Froelich, D., Hodson, B. W., & Edwards, H. T. (2002). Characteristics of Vietnamese phonology. *American Journal of Speech-Language Pathology, 11*(3), 264-273.

Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*(1), B47-B57.

Johnson, K., & Johnson, H. (1999). *Encyclopedic dictionary of applied linguistics: A handbook for language teaching*. Oxford: Blackwell Publishers.

Jones, R. H. (1997). Beyond "listen and repeat": Pronunciation teaching materials and theories of second language acquisition. *System, 25*(1), 103-112.

Kuhl, P. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America, 97*(22), 11850-11857.

Kuhl, P., Conboy, B., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions Of The Royal Society Of London. Series B, Biological Sciences, 363*(1493), 979-1000.

Matthews, P. H. (Ed.). (2007). *The concise Oxford dictionary of linguistics*. Oxford: Oxford University Press.

McCandliss, B. D., Fiez, J. A, Protopapas, A., Conway, M., & McClelland, J. L. (2002). Teaching the /r/-/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Cognitive, Affective, & Behavioral Neuroscience, 2*(2), 89–108.

Michas, I., & Berry, D. (1994). Implicit and explicit processing in a second-language learning task. *European Journal of Cognitive Psychology, 6*(4), 357-381.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289-306.

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research 71*(1). 4-12.

Nilsen, D. L., & Nilsen, A. P. (1973). *Pronunciation contrasts in English* (2nd ed.). New York: Regents.

Sheldon, A., & Strange, W. (1982) The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*(3), 243-261.

Tang, G. (2007). Cross-linguistic analysis of Vietnamese and English. *Journal of Southeast Asian American Education & Advancement,* 2, 1-33.

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*(3), 263-290.

Vita

Anna Lauzon was born in Houston, Texas on August 25, 1981. Her parents are Dr. and Mrs. J. P. Lauzon. Ms. Lauzon graduated from St. Agnes Academy in 2000 and received her Bachelor of Arts with Honors in English and a minor in Spanish from Davidson College in 2004. After teaching English in Spain and the United States for several years, she attended the University of Texas at Austin, where she received her Master of Arts in Foreign Language Education in 2009. While at UT, her collaborative research on the role of technology in teacher education and suprasegmentals in English as a Second Language research were presented at poster sessions during the Texas Foreign Language Education and American Association for Applied Linguistics Conferences.

Permanent address: 5019 Wigton Dr., Houston, TX 77096

This report was typed by the author.