7-12-2012 12:00 AM

Speech Compensation to Formant Perturbations in English and Vietnamese Talkers

Linh L.T. Nguyen
The University of Western Ontario

Supervisor
Dr. David W. Purcell
The University of Western Ontario

Graduate Program in Health and Rehabilitation Sciences
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

SPEECH COMPENSATION TO FORMANT PERTURBATIONS IN ENGLISH AND VIETNAMESE TALKERS

(Spine title: Speech Compensation to Formant Perturbations)

(Thesis Format: Monograph)

by

Linh Le Truc Nguyen

Graduate Program in Health and Rehabilitation Sciences

A thesis submitted in partial fulfillment
of the requirements for the degree of
Masters of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor	Examiners
Dr. David Purcell	Dr. Lisa Archibald
Supervisory Committee	Dr. Margaret Cheesman
Dr. Lisa Archibald	Dr. Jeffery Jones
Dr. Debra Jared	

The thesis by

Linh Le Truc Nguyen

entitled:

Speech compensation to formant perturbations in English and Vietnamese Talkers

is accepted in partial fulfillment of the
requirements for the degree of
Masters of Science

Date

Chair of the Thesis Examination Board
Dr. Janis Cardy

Abstract

The purpose of this experiment was to examine mechanisms underlying the auditory feedback system using Vietnamese and English talkers in response to feedback perturbations. F1 discrimination thresholds, vowel goodness ratings, and vowel category bounds for English /ɪ/ were determined. Vowel spaces were collected for both languages and auditory feedback of F1 was manipulated for English and Vietnamese vowels. Speech compensation during perturbed auditory feedback occurred in English and Vietnamese vowels suggesting that the underlying mechanisms are universal. However, there were differences in speech compensation for some vowel conditions, which may have occurred due to vowel location in each language group's vowel space. Speech compensation may also be influenced by the perceptual boundaries of internal vowel categories where significant compensation may occur when feedback is non-prototypical to the target vowel category. Overall, these results suggest that the feedback system is sensitive to formant frequency changes that may have underlying phonological mediations.

Index Terms: auditory feedback, language acquisition, speech production, speech perception, vowel formants, Vietnamese, Canadian-English

Acknowledgments

I would like to express my most sincere thanks to my research supervisor, Dr. David Purcell. Thank you for providing me with your positive encouragements, guidance, inspirations and many joyful memories over the past two years.

I would like to thank my advisory committee, Dr. Debra Jared and Dr. Lisa Archibald, for all the support and knowledge they have provided me. I would also like to give special thank you to Takashi Mitsuya and Dr. Ewen MacDonald for all your help.

I would like to give a special thank you to UWO Vietnamese student association and to all the participants who contributed to this body of work and without whom this study would not have been possible.

I am grateful to my SAFER labmates, friends and colleagues for their ongoing support and suggestions. Thanks especially to Teresa Valenzano and Kassandra Birtch for the moral support they've given, ears to listen and shoulders to lean on.

To my amazing fiancé, Daniel Vaccarello, for everything. You were always there when I needed your computer skills and your jokes to keep me calm. Thank you for being there every step of the way from applying to do my Masters to listening to my presentations endlessly.

Finally, I am especially grateful to my family for the unconditional love and support they have provided during the past two years, as always.

Funding for this project was provided by the University of Western Ontario, Early Researcher Award, and the Ontario Graduate Scholarship.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

2AFC	Two-alternative forced choice
µs	Microseconds
AOA	Age of arrival
BMO	Best model order
CPH	Critical Period Hypothesis
CPS	Common phonological space
dB	Decibels
ESL	English as a second language
F0	Fundamental frequency
F1	First formant
F2	Second formant
F3	Third formant
HL	Hearing level
Hz	Hertz
L1	First language
L2	Second language
LPC	Linear predictive coding
ms	Milliseconds
NRV	Natural Referent Vowel

OT-APST	Occupational-Therapy Adult Perceptual Screening Test
PCA	Phonetic category assimilation
PCD	Phonetic category dissimilation
SD	Standard deviation
SLM	Speech Learning Model
SPL	Sound pressure level
VOT	Voice onset time
yr	Years of age

Chapter 1

1 Introduction

Auditory feedback, hearing one's voice, may play a role in the detection of speech errors and regulation of speech production (Borden, Harris, & Raphael, 1994). If this system is perturbed, speakers will make changes in their speech to correct for the perturbation. Manipulations of the auditory feedback system have been investigated with voice pitch (Burnett, Freedland, Larson, & Hain, 1998; Jones & Munhall, 2000), loudness (Bauer, Mittal, Larson, & Hain, 2006; Lane & Tranel, 1971), and spectral characteristics of sounds (Garber, Seigel, & Pick, 1981). Speech compensations have also been measured with formant manipulated auditory feedback, where studies have perturbed the first formant (F1) and second formant (F2) (MacDonald, Goldberg, & Munhall, 2010; Munhall, MacDonald, Byrn, & Johnsrude, 2009; Purcell & Munhall, 2006ab; Villacorta, Perkell, & Guenther, 2007). This research has mainly used native English talkers, except for one study where they used Japanese and Korean talkers with English as a second language (ESL; Mitsuya, MacDonald, Purcell, & Munhall, 2010, 2011). A novel study has been conducted comparing English monolingual talkers with Vietnamese bilingual speakers with ESL using formant manipulations to determine the phonological influence on speech compensations. In the following introduction, a review of Canadian and Vietnamese vowels is provided, followed by a summary of language acquisition theories, auditory feedback, and speech perception research.

1.1 Introduction to Canadian-English vowels

There are ten vowels in the Canadian-English language: /i, e, ɪ, ɛ, æ, ɑ, ʌ, ʊ, o, u/ (Hagiwara, 2006). The Canadian-English vowel space can be seen in Figure 1.In contrast to other English dialects, the vowels /a/ and /ɔ/ have merged together in Canadian-English, which is called the Canadian Shift (Clarke, Elms, & Youssef, 1995). However, this Canadian Shift does not occur in all regions within Canada. The Canadian Shift has

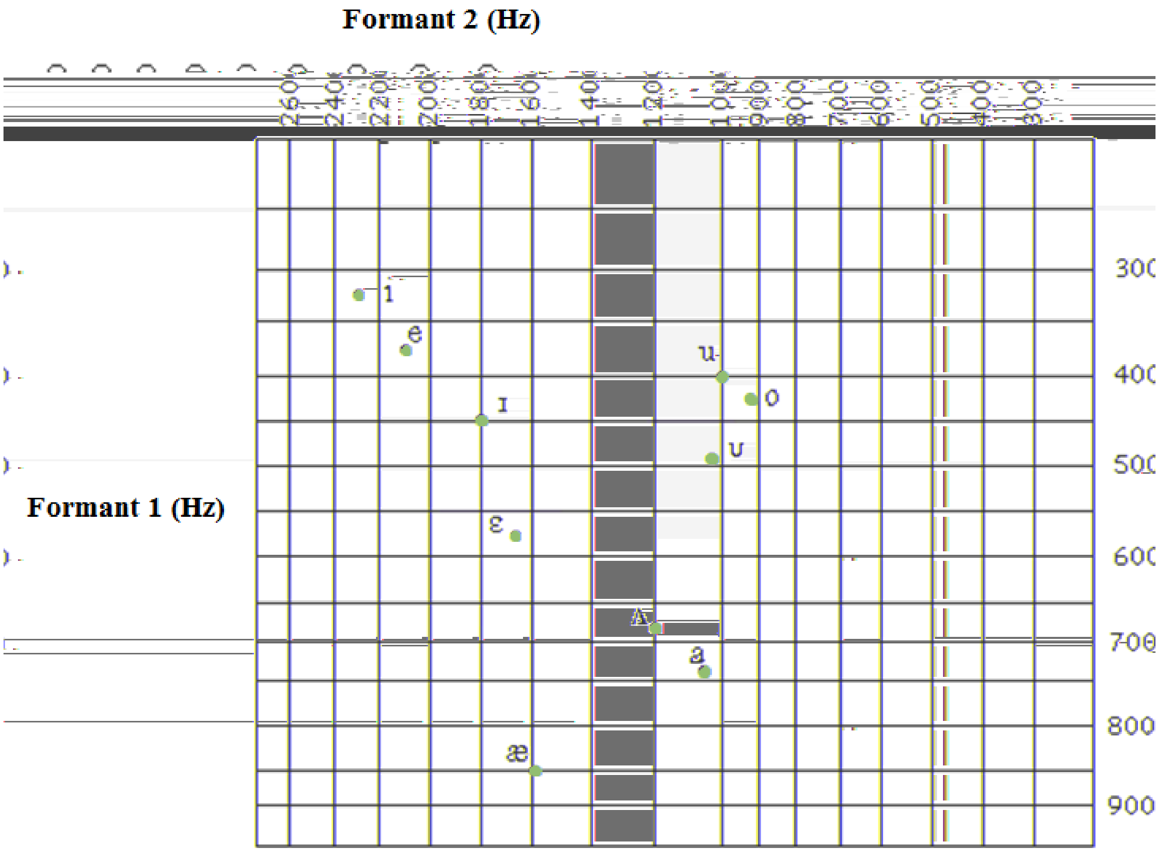


Figure 1: Canadian-English vowel space.

The vertical axis is F1 (Hz) and the horizontal axis is F2 (Hz). The plot was adapted from http://www.ic.arizona.edu/~lsp/Canadian/canphon2.html (Mendoz-Denton, Hendricks, & Kennedy, 2001).

been found to occur in Ontario (Clarke et al., 1995) and Winnipeg (Hagiwara, 2006). However, with Montreal-English there seems to be a shift in vowels that is different than the Canadian Shift (Boberg, 2005). As well, a combination of the Canadian and Montreal Shift occurs in St. John's-English (Hollett, 2006). Therefore, across Canada there are regional dialects that may differ slightly in their vowel spaces. Due to the variability in Canadian-English, individuals who learned English in the Maritimes and Quebec were excluded from the study.

1.2 Introduction to Vietnamese vowels

Vietnamese is an Austro-Asiatic language that is monosyllabic and tonal, with three dialects: Northern (Hanoi), Central (Hue), and Southern (Saigon) (Edmondson, 2009; Liem, 1970; Santry, 1997). There are six tones that are noted by diacritics placed above or under the vowel: level, falling, creaky, dipping-rising, rising, and constricted (Tang & Barlow, 2006). However, in the southern dialect there are only five tones where the creaky tone has merged with the dipping-rising tone (Tang & Barlow, 2006). There are 11 single vowel phonemes (also known as monopthongs or singletons): /i, e, ɛ, ɐ, a, ʌ, ɤ, ɔ, ɯ, o, u/ (see Figure 2; Chen, Li, Shen, & Fu, 2001; Liem, 1970; Santry, 1997; Tang & Barlow, 2006). Vietnamese dialects differ mostly on consonant sounds rather than vowel production (Tang & Barlow, 2006). Since the current study uses Vietnamese vowels only, Vietnamese talkers of different dialects are analyzed as one group. There is greater density in the mid-low, front-central Vietnamese vowel space area (Figure 2) than in the same area of the Canadian-English vowel space (Figure 1).

1.3 Acquisitions and interactions between first (L1) and second (L2) language

There are many differences between monolingual native (L1) speakers and those who speak the L1 as their second language (L2). Usually this difference is detectable through an accent spoken by the L2 speaker and an error in the perception of the L2 words, vowels, and consonant sounds. However, some L2 speakers resemble L1 speakers while

A)

		Front	Central	Back	
				Unrounded	Rounded
High		i		ɯ	u
Mid	Close	e		ɤ	o
	Open	ɛ	ʌ		
Low	Close		ɐ		ɔ
	Open		a		

B)

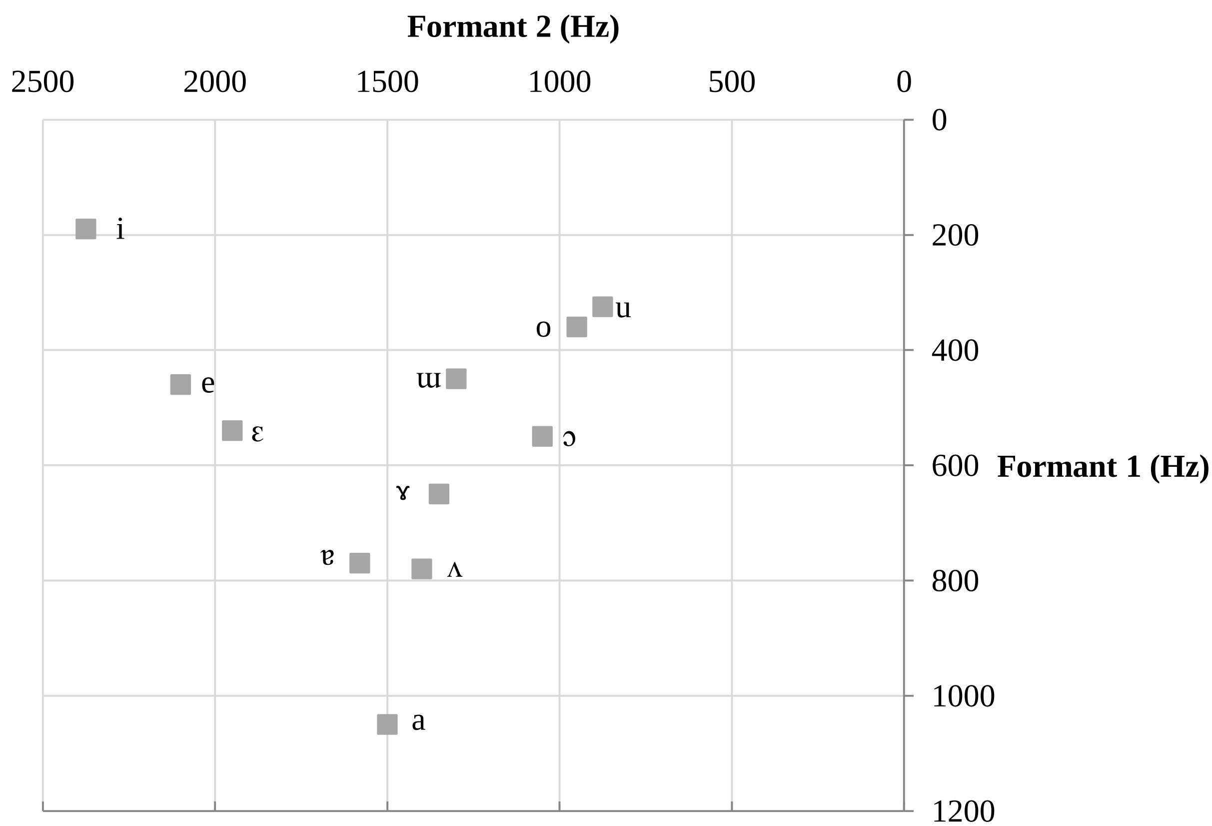


Figure 2: Vietnamese vowels.

A) Vowel chart is a replication as seen in Liem (1970, p. 12; refer also to Santry (1997, p. 17). B) Vietnamese vowel space. The vertical axis is F1 (Hz) and the horizontal axis is F2 (Hz). F1/F2 frequency values were provided by Santry (1997, p. 147).

others are very different. These differences between the L2 speakers are dependent on various factors, such as the age when they acquired the L2, the amount of L1 and L2 usage, and interactions between the L1 and the L2.

1.3.1 Critical Period Hypothesis (CPH): The effects of age

There is consensus in the literature that individuals who learn the L2 during childhood (early learners/bilinguals/speakers) are more similar to native speakers of L2 than individuals who learn the L2 during late adolescence or adulthood (late learners/bilinguals/speakers). A study by Flege, Yeni-Komshian, and Liu (1999b) has found that early learners have similar sensitivity to grammatical properties as native speakers. As well, early bilinguals are able to online process L2 sentences like native speakers (Guillelmon & Grosjean, 2001; Liu, Bates, & Li, 1992; McDonald, 2000). In addition, early bilinguals have mild or no accents in the L2 when compared with late bilinguals (Flege, Munro, & MacKay, 1995a; Flege et al., 1999b). Furthermore, early speakers are able to perceive and produce the L2 consonants and vowels like native speakers (Flege, 1991, 1992; Flege, MacKay, & Meador, 1999a; Flege, Munro, & MacKay, 1995b; MacKay, Flege, Piske, & Schirru, 2001a; MacKay, Meador, & Flege, 2001b; Munro, Flege, & MacKay, 1996; Piske, Flege, MacKay, & Meador, 2002). Finally, Meador, Flege, and MacKay (2000) showed that early learners are able to more accurately recognize L2 words in noise than late learners. Overall, to resemble the native speakers of a specific language, the individual should learn the language at a young age.

Through these observations, it has been proposed that L2 acquisition is constrained by a critical period (Johnson & Newport, 1989; Lenneberg, 1967; McLaughin, 1977; Patkowski, 1990). The Critical Period Hypothesis (CPH; Lenneberg, 1967) suggests that the capacity to learn another language after the critical period declines and learning becomes difficult. This critical period has been suggested to be around the age when the individual reaches puberty (Lenneberg, 1967), at 12 years of age (yr; Scovel, 1988), 15 yr (Patkowski, 1980, 1990), at 6 yr to be accent-free in the L2 (Long, 1990), or between the ages of 6-7 or 16-17 (DeKeyser, 2000). Furthermore, both Seliger (1978) and Walsh and Diller (1981) suggested that there are critical periods corresponding to each aspect of language learning that are closed off at different times, and the ability to have a native accent in the foreign language is lost first around puberty. Due to the age variances for the critical period, it is also known as the sensitive period (Long, 1990; Oyama, 1976).

There are developmental effects or maturational constraints that occur after the sensitive period that causes a decline in language acquisition. After the sensitive period, the mechanisms that are used for successful language acquisition become inaccessible for use or work less effectively (Flege, 2002). For instance, Scovel (1988) suggested that after 12 years, neurological maturation in the central nervous system causes a decrease in brain plasticity from a reduction in myelination, in which circuits that are used are reinforced and circuits that are not used would weaken; as a result difficulty in learning ensues. In contrast, Bever (1981) suggested that before the critical period, psychogrammar, a mechanism that is used to simultaneously learn the production and perception of a language, is available. After the critical period, psychogrammar decays, marking the end for speech learning. When an individual is learning a language after the critical period, production and perception develop independently. As a result, synchronicity issues arise between language production and perception, such that an individual learn to discriminate sounds but cannot produce them (Bever, 1981). Therefore, the reduction in brain plasticity and the decay of psychogrammar after puberty cause L2 learning to become difficult.

The CPH should be held with caution because various studies have suggested that late learners are able to learn the L2 like L2 native speakers and early learners may not be as proficient in the L2 as L2 native speakers (Birdsong, 1992; Bongaerts, van Summeren, Planken, & Schils, 1997; White & Genesee, 1996). A study by Neufeld (1979) was able to show that adult L2 learners were able to access the mechanisms involved for language acquisition that are supposedly inaccessible after puberty. In his study, he trained 20 English-speaking students to imitate ten short phrases in Chinese and Japanese with an 18 hour program. Native speakers of each language found that one subject had a native accent in both languages, and two subjects had native accents for Japanese only. Therefore, the perceptual and motor mechanisms involved in speech production are accessible after puberty (Bongaerts et al., 1997). Furthermore, Scovel (1988) suggested that there are “superexceptional” learners who are late bilinguals that resemble L2 native speakers. According to him, about one in 1000 late L2 learners are not bound by the critical period constraints. For instance, Schneiderman and Desmarais (1988) had two native English speakers who had acquired French during adulthood, yet were judged to come from French-speaking communities by four native Francophones. As well, one of the English speakers had also acquired Spanish and was judged to sound like a native Spanish speaker. Similar results were found by Ioup (1995), Ioup, Boustagui, El Tigi, and Moselle (1994), and Novoa, Fein, and Obler (1988). Therefore, the CPH provides strong predictive value in determining the success of an individual learning a foreign language, but it is not always valid and reliable.

1.3.2 The amount of first (L1) and second (L2) language usage

A bilingual’s L1 might be their native language; however, it might not be their dominant language. A dominant language is the language that is used the most by the speaker. The L2 would become a person’s dominant language if the majority of their everyday activities involve the L2; as a result, the L2 would be used more often and be developed to a greater extent than the L1 (Grosjean, 1982). Therefore, a person’s abilities in their L1 and L2 are affected by the contexts in which the L1 or L2 are used.

The amount of L1 and L2 exposure and practice are inversely correlated with, and dependent on, age of arrival (AOA) in the L2 country. A study conducted by Flege et al. (1995a) looked at 240 Italian immigrants to Canada, with AOA from 2-22 years of age and living in Ottawa for about 32 years. The authors found that the Italians who came to Canada during early childhood used English about four times more often than Italian. In comparison, those who came during their young adulthood reported using English and Italian nearly equally. Similarly, Flege et al. (1999b) examined 240 Korean immigrants to the United States and reported the ratio of English to Korean usage. They found that the immigrants who arrived to the United States between 2 and 12 years of age had English dominance; however, those who arrived between 13 to 23 years of age used both English and Korean equally. A re-analysis of the data collected by Yeni-Komshian, Flege, and Liu (2000) was performed by Flege (2002). In the Yeni-Komshian et al. (2000) study, Korean-English bilingual participants reported their proficiency in Korean and English, in terms of their pronunciations and memory of word pronunciation, grammar knowledge in both languages, and abilities to write and read. Flege (2002) found that immigrants who came before 8 yr (early bilinguals) were English dominant. In contrast, immigrants who came after 16 yr (late bilinguals) were Korean dominant. The general consensus is that early bilinguals will be dominant in their L2 because of their daily surroundings of high L2 use. These authors suggest that late bilinguals will be dominant in their L1 because of limited possibilities for using L2. The L2 learning would then be adversely affected for the late bilinguals as they continue to use their L1 as their dominant language.

1.3.3 Interactions between first (L1) and second (L2) language

1.3.3.1 Cross-language interference

Differences between native and non-native speakers in L2 performance were thought to be from cross-language “interference” (Flege, 2002). Interference is defined as the effect of previous learning on subsequent learning (Flege, 2002). According to Grosjean (1982), language interference can also include the effect of subsequent learning on previous learning. Therefore, there is bi-directional interference in which the dominant language will affect the non-dominant language.

This interference appears in language dominance for early and late bilinguals. Grosjean (1982, 1989, 1997; Grosjean & Soares, 1986) suggested that for early bilinguals, there would be a greater influence of the L2 on the L1 because the L2 is dominant. Whereas, for late bilinguals, there would be a greater influence of the L1 on the L2 because the L1 is dominant. He also stated that neither the L1 nor the L2 could be fully suppressed or deactivated. As a result, bilinguals would always be different from monolinguals.

Evidence for interference effects have been shown through detection of foreign accents. The study by Yeni-Komshian et al. (2000) studied 240 Korean immigrants arriving in the United States at various AOA. The participants repeated English and Korean sentences, and native English and Korean speakers were asked to rate the foreign accents of the participants. As controls, monolingual Korean and English speakers repeated sentences in their own respective language. The English sentences spoken by the bilinguals had lower ratings, indicating stronger foreign accents, than English monolinguals. Foreign accents were also heard in the bilinguals for spoken Korean sentences in comparison to Korean monolinguals. Because all the bilinguals were shown to have foreign accents, these results support Grosjean's theories of bilinguals being unable to completely deactivate one language. Furthermore, the relationship between having foreign accents and AOA was negatively correlated: early bilinguals would have stronger English accents in spoken Korean; whereas, late bilinguals would have stronger Korean accents in spoken English. Overall, foreign accents are detected in both languages for bilinguals, indicating interference.

1.3.3.2 Speech Learning Model (SLM)

The Speech Learning Model (SLM) was proposed by Flege (1995) to provide an explanation for changes in language acquisition and learning across the speaker's lifespan (Sebastian-Galles & Bosh, 2005). SLM maintains that the mechanisms and processes involved during language acquisition in childhood remain intact and accessible during adulthood for L2 learning. The inabilities of the late bilinguals to learn as well as early bilinguals are due to other factors, such as the learning environment and the regularity of interacting with L2 native speakers. The differences that arise from L2 learners and L2 native speakers occur because L2 learners continue to use their L1, causing L1 to influence L2. L1 and L2 can influence each other because there is a "common phonological space" (CPS, also called the L1-L2 phonetic space). In the CPS, all the phonic elements (vowels and consonants) of both languages are present and interacting with each other. Two types of interactions can occur with L2 learning: phonetic category assimilation (PCA) or phonetic category dissimilation (PCD). The type of interactions that occur between the L1 and the L2 are dependent on the similarities or differences of the specific L2 phonic element to the closest L1 phonic element.

When learning a native language, an individual will create a phonological space that contains only L1 phonic elements. Each L1 phonic element will have its variants;

however, all the variants of the specific L1 phonic element will be judged to be the same element, creating a category for the specific sound. This L1 phonetic space develops during childhood and through adolescence as the individual is mastering their native language. When L2 learning occurs, the L2 phonic elements would be added into the L1 phonetic space creating the CPS. Each instance of an L2 phonic element will be compared to the closest L1 phonetic category and PCA or PCD will occur.

1.3.3.2.1 Phonetic category dissimilation (PCD)

PCD will occur when the perceived phonetic distance between the L2 phonic sound and the closest L1 phonetic category is large. When this distance is large, the L2 learner will create a new category in the L1-L2 phonetic space for the specific L2 sound. As new categories are formed in the L1-L2 phonetic space, the space becomes fuller. To maintain contrast between neighbouring categories, the categories may deflect away from each other (Bohn & Flege, 1992). The distance between the established L2 category and the nearest L1 category would increase to maintain its uniqueness. As a result, the phonetic categories of bilinguals would not be the same as the phonetic categories of L1 or L2 monolinguals, implying that bilinguals will always have a foreign accent in the L1 and L2 in comparison to their monolingual counterparts.

PCD has been shown to occur in studies with bilingual children and adults. Flege and Eefting (1987) examined Spanish children and adults in Puerto Rico, where English was taught in a bilingual school (Spanish-English bilinguals). All the participants were considered to be early bilinguals because English was learned first during childhood. The controls in the experiments were monolingual Spanish and English speakers who were age-matched to the bilinguals. Participants were asked to read Spanish and English words. The researchers measured the voice onset time (VOT) of /p, t, k/ in English and Spanish because these tokens have long-lag VOT in English and short-lag VOT in Spanish. Among the monolinguals, there were differences in VOT between the adults and children within each language. This supports the SLM’s hypothesis that the categories in the phonological space mature slowly over time. Results from the Spanish-English bilinguals showed that the bilinguals produced longer VOT for /p, t, k/ in English than their age-matched English monolinguals; in contrast, the bilinguals produced shorter VOT for /p, t, k/ than their age-matched Spanish monolinguals. These results lend support to the PCD because the bilinguals showed an exaggeration of the monolinguals' VOT for /p, t, k/. This indicated that new categories were formed for /p, t, k/ in the phonetic space when English was learned. To make the Spanish and English categories distinct, the categories deflected away from each other causing VOTs to be longer than English monolinguals and shorter than Spanish monolinguals. Other studies such as Flege, Schirru, and MacKay (2003) and Mack (1990) have also demonstrated PCD to occur during L2 acquisition.

1.3.3.2.2 Phonetic category assimilation (PCA)

PCA will occur when the perceived phonetic distance between the L2 phonic element and the closest L1 phonetic category is small. When this distance is small, the individual will always judge the L2 phonetic sound to be instances of an established L1 phonetic category. This process is called equivalence classification: the specific L2 phonetic sound is said to be "equated" to the established L1 phonetic sound and the L2 phonetic sound is assimilated into the specified L1 phonetic category. Therefore, new category formation has been prevented or blocked, and as a result, the speaker will use the corresponding L1 phonetic element to produce and perceive the specific L2 phonetic sound.

PCA has been shown to occur in studies observing the perception and production of stop consonants. Flege (1987) examined two groups of bilinguals: 1) American women who were residents of Paris for 12 years and learned French during adulthood (English-French bilinguals) 2) French or Belgian women who were residents of Chicago for 12 years and learned English during adulthood (French-English bilinguals). The control groups were English and French monolinguals. The participants were asked to read French and English sentences, and the VOT for /t/ tokens were determined. English /t/ tokens were spoken with a long-lag VOT; in contrast, French /t/ tokens were spoken with a short-lag VOT by their respective monolinguals. Results showed that both groups of bilinguals had differences in VOT for /t/ when they spoke in English or French, indicating their tacit knowledge that there is a difference between the languages. However, these differences were not as large as the VOT by the monolinguals. Furthermore, English-French bilinguals produced the French /t/ tokens with longer VOTs than French monolinguals, indicating an influence of the English /t/; whereas, the French-English bilinguals produced the English /t/ tokens with shorter VOTs than English monolinguals, showing an influence of the French /t/. These results provided evidence for PCA because it showed that category formation for the L2 /t/ was blocked, and instead the L1 and L2 categories for /t/ were merged. As a result of this merged category, differences between the VOT for /t/ occurred between the monolinguals and bilinguals.

PCA and PCD occurred in late and early Italian-English bilinguals in the study by MacKay et al. (2001a). These results suggested that PCA occurs throughout a person's lifespan. The PCA mechanism is more influential than PCD because as the L1 phonetic space develops and matures during childhood and adolescence, the L1 phonetic categories become stronger attractors for the new L2 phonic elements. A study by Flege et al. (2003) was able to demonstrate the likelihood of PCA or PCD to occur across a lifespan. In their study, Italian immigrants with varying AOA to Canada (early or late Italian-English bilinguals) participated. These participants were asked to produce instances of /e/ in Italian or English, where the English /e^{I}/ is produced with more tongue movements than the Italian /e/. The early bilinguals produced the English /e^{I}/ with more exaggerated tongue movements than English monolinguals, indicating that PCD had occurred. PCD occurred because a new category was formed for the English /e^{I}/. As a result, to differentiate between the Italian /e/ and English /e^{I}/, the categories deflected, exaggerating the tongue movements for the English /e^{I}/. In contrast, late bilinguals produced the English /e^{I}/ with less tongue movements than English monolinguals, indicating that PCA had occurred. PCA occurred in the late bilinguals because the production of the English /e^{I}/ was more similar to the Italian /e/. Therefore, there is greater likelihood for new L2 phonic elements to assimilate into an established L1 category as the L1 phonetic space develops and matures.

1.4 Auditory feedback

The detection of speech errors and regulation of speech production may be monitored by auditory feedback, the hearing of one's voice. If the auditory feedback is perturbed, the speaker will make changes in their speech to correct for the perturbation. For instance, if the heard sound is amplified, speakers decrease their vocal intensity, whereas if the sound is attenuated, speakers increase their vocal intensity (Borden et al., 1994). As well, if the sound is filtered, speakers make speech adjustments to modify some of their vocal tract resonance characteristics so that the speech sound target could be achieved (Garber et al., 1981). Speech compensation effects have also been found in pitch-shifted studies that perturbed the fundamental frequency (F0) in English talkers. Talkers compensate in the opposite direction of the F0 perturbation (Burnett et al., 1998; Jones & Munhall, 2000). Jones and Munhall (2002) also replicated these results in Mandarin talkers. Therefore, a general pattern has emerged across studies where the auditory feedback was manipulated: speakers change their speech production in the opposite direction of the perturbation.

This general pattern has also been found in studies in which the auditory feedback of normal hearing individuals has been manipulated by altering vowel formants and measuring changes in speech production (MacDonald et al., 2010; Mitsuya et al., 2011; Munhall et al., 2009; Purcell & Munhall, 2006ab). This paradigm generally involves four stages in which the speaker is producing vowels: Baseline, Ramp, Hold, and End, as seen in Mitsuya et al. (2011; see Figure 3). In the Ramp phase, two directions of the formant perturbation are often used, where a given formant is either shifted up or down. For example, if F1 of the vowel /ɛ/ in 'head' is decreased, it would slowly sound like /ɪ/ in 'hid'; if F1 is increased, it would slowly sound like /æ/ in 'had'. The rate at which the perturbation is introduced does not affect compensatory responses, although they are dependent on the magnitude of the perturbation. MacDonald and colleagues (2010) examined the effect of different magnitudes of formant frequency changes in speech compensation. In their first experiment, participants received a three-step change in acoustic feedback: F1 was shifted up by 50, 100, and 200 Hertz (Hz) and F2 was shifted

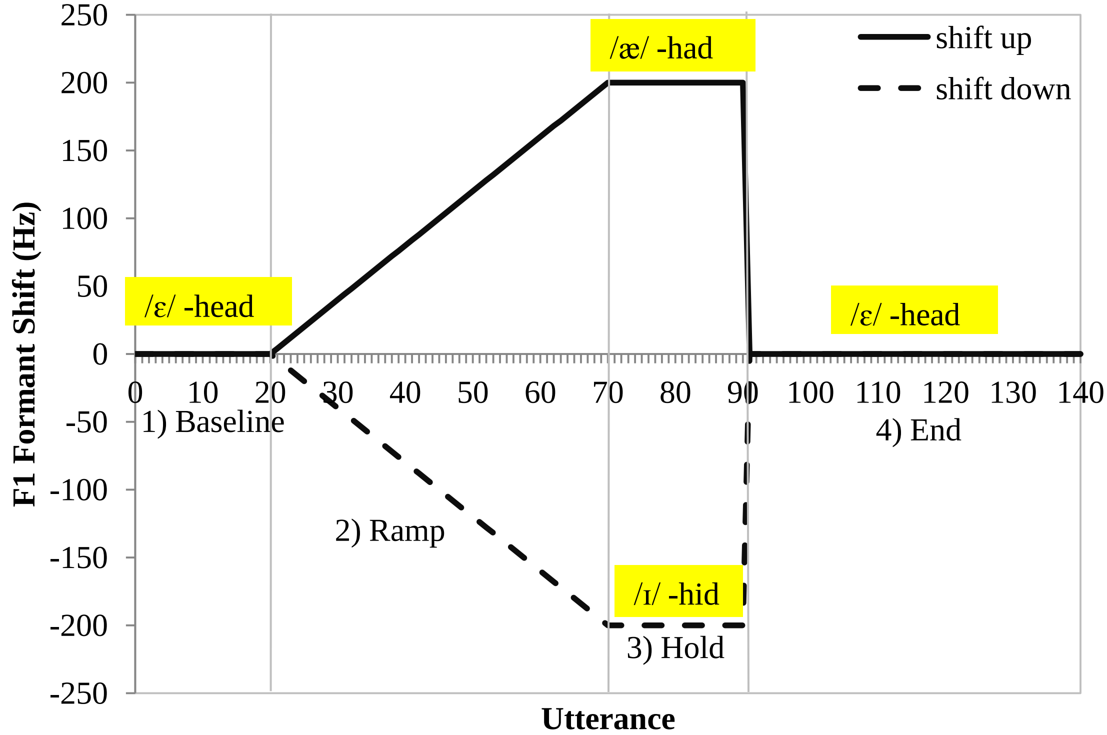


Figure 3: Schematic diagram of the phases for the formant shifts.

The diagram was adapted from Figure 1 in Mitsuya et al. (2011).

down by 75, 125, and 250 Hz. In their second experiment, the acoustic feedback was shifted gradually: F1 was increased by +4 Hz, and F2 was decreased by -5 Hz on each utterance. Results found that the compensatory responses were similar at each magnitude change, when compared between the experiments. The magnitude of F1 change in formant manipulated auditory feedback is usually ± 200 Hz because this shift would change the sound of the target vowel to another vowel category.

1.4.1 Responses to formant perturbations

Speech compensations by the motor control system can adjust F1 and F2 independently. Studies such as MacDonald et al. (2010) and Munhall et al. (2009) have manipulated F1 and F2 concurrently. The advantage of shifting F1 and F2 concurrently allows for a natural vowel sound. However, a disadvantage is that estimates of some talkers' F2 values are unstable, leading to unnatural transients in the feedback. Studies have shown that the system is able to correct for F1 and F2 perturbations relatively independently, as a result, manipulating two formants concurrently is not always required. A study by Villacorta et al. (2007) examined the speech production of F1 and F2 while shifting F1 gradually. Their results found a compensation for F1 while F2 productions remained stable. In addition, MacDonald, Purcell, and Munhall (2011) examined the ability of the motor control system to correct for F1 and F2 independently. They found that a perturbation in a single formant did not require the system to change the entire vowel spectrum but only the manipulated formant. Overall, the manipulation of a single formant results in compensation of the manipulated formant.

The perturbed feedback elicits speech compensation that is proportional to the manipulation. Talkers in formant manipulated studies usually compensate at ~25%-50% of the formant perturbation (Houde & Hordan, 1998; Liu & Larson, 2007; MacDonald et al., 2010; Munhall et al., 2009; Purcell & Munhall, 2006a; Villacorta, 2007). Partial compensation reflects that speech is regulated by systems other than audition, such as the somatosensory system and the feedforward system (Nasir & Ostry, 2008; Tremblay, Schiller, & Ostry, 2003). There are a few possibilities that may explain the partial compensation. One possibility is that for some vowels, there are physical constraints that may prevent full compensation. For instance, compensation for /i/ in the positive shift would be limited because the tongue movements required would be higher than the palate position (MacDonald et al., 2010). Due to these physical differences in pronouncing vowels, the system may weigh the importance of auditory feedback and somatosensory feedback differently for each vowel. For instance, /i/ production could rely more on somatosensory information than auditory feedback and as a result partial compensation may occur. For these reasons, generally mid-vowel /ε/ has been used to limit available tactile feedback (MacDonald et al., 2010; Mitsuya et al., 2011; Purcell & Munhall, 2006a). The magnitude of the perturbation may also cause partial compensation (MacDonald et al., 2010). If the magnitude of the perturbation is too large, the auditory feedback system may ignore the feedback and attribute it to other environmental sources or treat it as unrealistic. Due to these possibilities, partial compensations are expected in altered auditory feedback studies.

Compensation was also found to be automatic and to occur unconsciously. A study by Munhall et al. (2009) used different instructional conditions with the sudden large perturbation paradigm similar to MacDonald et al. (2010). The instructional conditions were: (1) *control*: naive subjects that had no information about the feedback manipulation, (2) *ignore headphones*: the subjects were made aware of the changes occurring in the headphones and were instructed to ignore them, and (3) *avoid compensation*: subjects were made aware of the manipulation and were instructed to maintain normal speech production without compensation. All three instructional groups altered their vowel formant values in the opposite direction to the perturbation and there were no differences in the magnitude of these formant changes. Houde and Jordan (2002) also found in post-experiment interviews that participants were not aware of the feedback manipulation or compensation responses. This suggests that the participants were not using conscious strategies to compensate for the perturbed feedback. Compensation behaviour does not change with conscious strategies because these studies have shown it occurs automatically.

1.4.2 Phonological mediation in speech compensation

Many of these studies investigating the role of auditory feedback in speech production have used participants with English as L1. The use of auditory feedback may differ when a person is using their native or secondary language. A preliminary cross-language study by Mitsuya et al. (2010) examined the use of auditory feedback in shaping articulation patterns for English, Japanese, and Korean vowels with Japanese and Korean ESL talkers using a paradigm similar to Purcell and Munhall (2006b). The experimental manipulation was to either shift F1 up or down for the English vowel /ɛ/ in ‘head’, a Japanese vowel /e̞/, and a Korean vowel /ɛ/. In general the talkers, when speaking in English or in their native language, compensated in the opposite direction of the perturbation. This indicated that the role of auditory feedback in speech motor control appears to be similar across languages. However, when the groups of talkers were compared for compensations for English /ɛ/, group differences between native English talkers and the ESL talkers occurred. The native English talkers had significant compensations earlier during the ramp phase (the speech compensation threshold) than Japanese talkers. This suggested that native English speakers were more sensitive to feedback error and the feedback system is able to correct speech production if there are small errors in auditory feedback. They also found that the Japanese and Korean speakers had smaller compensations compared to the native English speakers. As well, the stability of the compensatory behaviour of English speakers was greater than that of the ESL groups. These results suggest that in a secondary language, the sensitivity to feedback error is reduced and there is more variable production.

The results in the Mitsuya et al. (2010) study may have occurred due to the differences of the vowel spaces in English, Japanese, and Korean. The native English speakers may have produced larger compensations than the Japanese and Korean ESL speakers because the native English speakers have had more experience with English vowels. With this experience, more precise vowel boundaries may be defined, and as a result the native English speakers might be more sensitive to the perturbations when formant frequencies are changed. Also, when the Japanese and Korean ESL groups are learning the English /ɛ/ sound, they may have assimilated the English /ɛ/ into their own native language vowel space. As a result, the ESL talkers may use their native vowel space in response to the altered feedback. Mitsuya et al. (2010) reported that the density around /ɛ/ in the English vowel space is greater than in the same area of the Japanese and Korean vowel spaces. They suggested that native English talkers had greater compensatory behaviour than ESL speakers because smaller perturbations in /ɛ/ could sound like another English vowel. In contrast, the /ɛ/ vowel could have a larger space in the ESL talkers' vowel space because there are fewer vowels around that area; thus, it might require a larger magnitude of formant perturbations for the ESL speakers to respond to the change. Therefore, the perception of the feedback error may influence the compensation behaviours of talkers. If the feedback is perceived to be an acceptable token, there would be less compensation to maintain the perceptual distinctiveness of the produced vowel (Mitsuya et al., 2011). This suggests that acoustic feedback is not a purely frequency-based error reduction process but is also phonologically mediated.

If the density of the vowel space is supposed to affect compensation, results should indicate that denser vowel spaces elicit greater compensation to maintain vowel category distinctiveness. However, studies have found that distances to neighbouring vowel categories do not affect the magnitude of compensation. MacDonald et al. (2011) determined the neighbouring vowel distances of /ɛ/ to /ɪ/ and /æ/, which corresponded to the negative and positive shifts of /ɛ/, respectively. The correlations between compensation during the shifts and the respective vowel distances to the next vowel category were not significant. MacDonald et al. (2010) also had similar findings. These results suggest that compensation is not affected by the distances of neighbouring vowels and that auditory feedback is reacting to differences in intended and received frequencies. Villacorta et al. (2007) also supports this as they found that participants with lower F1 discrimination thresholds had greater compensation to the F1 feedback perturbations. These studies suggest that the control system is acting on the target vowel category only and neighbouring vowel categories do not influence the system.

However, others studies such as Mitsuya et al. (2011) and Purcell, MacDonald, and Munhall (2011) suggest that auditory feedback is partially phonologically mediated. Mitsuya et al. (2011) compared the negative and positive compensations for English /ɛ/ in

Japanese ESL and native English talkers. English /ɛ/ acts as a central vowel in English talkers and perturbations would occur within the English vowel space, creating similar compensations in the positive and negative directions. In contrast English /ɛ/, when perturbed positively in Japanese talkers, is outside the Japanese vowel space and there was less compensation in Japanese talkers than English talkers. However, when English /ɛ/ was perturbed negatively in Japanese talkers, the perturbations occurred within the Japanese vowel space and compensations were similar in Japanese and English talkers. Mitsuya and colleagues (2011) suggest an influence of vowel location on compensatory responses. Similarly, Purcell et al. (2011) found compensation differences across six English vowels. Purcell et al. (2011) found that point vowels such as /i, u/ in 'heed, who'd' had smaller compensations than more central vowels such as /ɛ/ in 'head'. As well, they found an asymmetry in speech compensations between the positive and negative shifts in some vowels like /i, ɪ / in 'heed, hid'. These differences in compensations may have been mediated by the location differences in the vowel space. Furthermore, that auditory feedback is partially phonologically mediated is supported by language acquisition and speech perceptual theories. Language acquisition theories, such as cross-language interference (see section 1.3.3.1) and the SLM (see section 1.3.3.2) suggest phonemes are interacting with each other creating differences between bilingual and native talkers. Speech perception theories such as category goodness (see section 1.5.1) and the Natural Referent Vowel (NRV) perspective (see section 1.5.2) also suggest vowel categories influence the participant's ability to discriminate and categorize. Further investigation is needed to explore the phonological influence on the auditory feedback system.

1.5 Speech perception

Vowel representations such as category prototypes, category goodness dimensions, and vowel location in perceptual space, may influence compensation to altered auditory feedback (Mitsuya et al., 2011). It is important to have an understanding of these phonological measures to determine their role in auditory feedback.

1.5.1 Vowel category goodness and bounds

Categories are created in perceptual systems to group similar stimuli together. However, all members of a category have varying degrees of category goodness, where some members are perceived as good exemplars (prototypical) or bad exemplars (Rosh, 1975). Members of a category are rated differently by listeners in terms of their category goodness (Grieser & Kuhl, 1989; Kuhl, 1991, 1986; Miller & Volaitis, 1989). Kuhl (1991) asked listeners to provide category goodness rating to /i/-like sounds based on their own internal prototype of /i/. Kuhl found that listeners consistently gave the experiment's prototype /i/ the highest category goodness rating and gave lower category goodness ratings to exemplars moving away from the prototype /i/. Therefore, internal categories are consistent across individuals and the exemplars in the categories are not all perceptually equal.

Category bound, the boundary of a good and bad prototype of a vowel, may be important in speech compensation to perturbed auditory feedback. Studies such as MacDonald et al. (2010) and Mitsuya et al. (2011) collected talkers' vowel spaces, where average F1/F2 frequencies of the vowels are plotted with ellipses representing one and two standard deviations (see Figure 4). These ellipses represent the production variability of the respective vowel and may not represent the category boundary of the vowel (Mitsuya et al., 2011). It is important to determine the category boundary of the vowel because significant speech compensation may start near this category bound. Before the category bound, feedback may be perceived as good exemplars of the target vowel; however, these exemplars are not identical to the prototype and small compensations may occur. After the category bound, feedback may be perceived as bad exemplars of the target vowel and significant speech compensation may occur. Further investigation is needed to determine the influence of category goodness and bounds on speech compensation during altered auditory feedback.

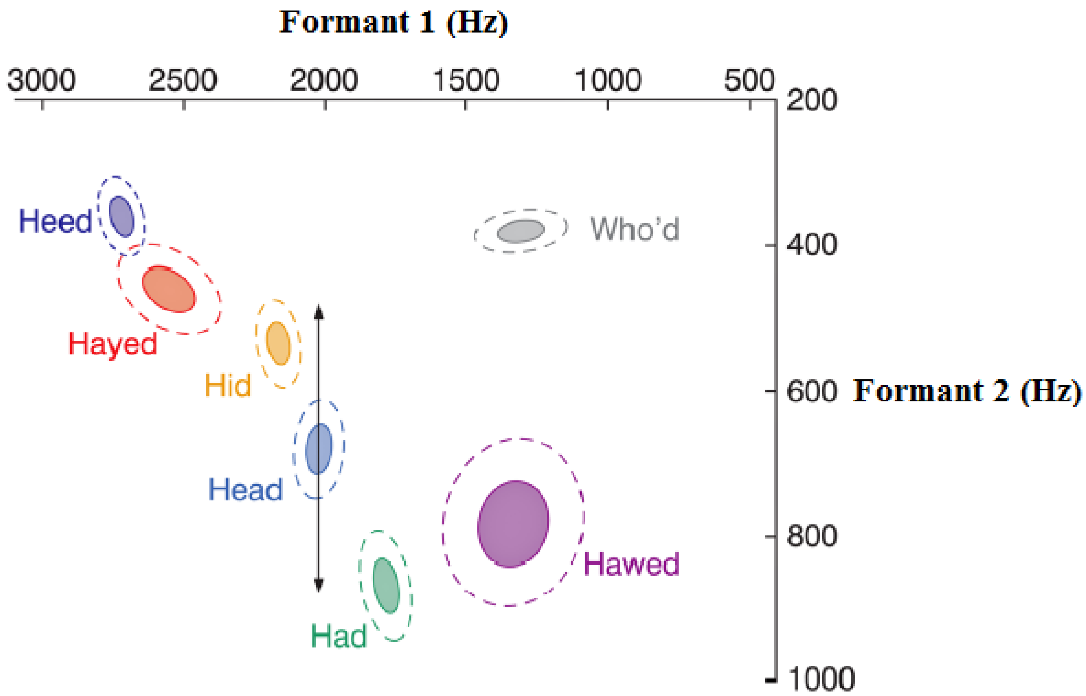


Figure 4: Vowel space of an example individual native English speaker.

The center of each ellipse represents the mean F1/F2 of the respective vowel in an /hVd/ context. The solid and dashed ellipses represent one and two standard deviations, respectively. Vowel space was adapted from Mitsuya et al. (2011) from Figure 6a.

1.5.2 Natural Referent Vowel (NRV) perspective

In vowel perceptual tasks, there are asymmetries in discrimination, where the direction of a change within the vowel space plays a key role. Direction of a vowel change is defined as the continuum of a vowel sound change, such as /ε → æ/ or /æ → ε/. One direction of a vowel change is easier to detect than the reverse direction with the same change. This suggests that vowels in a phonological space are not equally salient or familiar. The NRV perspective (also known as the peripherality hypothesis) suggests that vowels that are more peripheral (towards the corners of the vowel space) are more salient than vowels that are less peripheral (Polka & Bohn, 2003). These corner vowels act as a reference or perceptual anchor. In perceptual tasks, in which the direction is moving towards a peripheral vowel (i.e. /i, æ, α, u/), discrimination would be easier than moving away from a peripheral vowel. For instance, in the study by Polka and Bohn (1996), they had infants discriminate between English /æ/ and /ε/, in two directions /ε → æ/ and /æ → ε/. Results showed that discrimination was easier if the direction of the perturbation occurred from /ε → æ/. A study by Swoboda, Kass, Moose, and Leavitt (1978) has also found it was easier to discriminate /ɪ/ and /i/ if the direction was /ɪ →i/. This pattern of asymmetry in vowel perception has also been found to occur in German (Bohn & Polka, 2001; Polka & Bohn, 1996; Polka & Werker, 1994), cats (Heinz, Alesczyk, & May, 1996), and blackbirds (Hienz, Sachs, & Sinnot, 1981). Therefore, directional asymmetry in vowel perception seems to be language-universal and is not species-specific.

The direction of the perturbed feedback shift may have an effect on the speech compensation in perturbed auditory feedback. When the feedback is manipulated such that the perturbation is going towards a peripheral reference vowel, the auditory feedback system may be able to more easily discriminate between intended and perceived feedback, which may lead to greater compensation. The results of Mitsuya et al. (2011) may have been influenced by the NRV. The positive manipulation of /ε/ was towards the peripheral reference vowel /æ/ in English talkers which may have led to greater compensation. However, the positive shift elicited less compensation in Japanese talkers, perhaps because the shift was not towards a referent vowel. Vowel location may influence compensation results because direction of the perturbation in the feedback may provide an advantage for the auditory system to discriminate errors.

1.6 Rationale

The purpose of the present experiment was to determine if the perceptual organization of L1 and L2 vowels influences Vietnamese and English speakers' compensatory behaviour during speech production when auditory feedback is manipulated. To my knowledge, the studies by Mitsuya et al. (2010, 2011) are the only cross-language formant perturbation studies that have been reported in the literature. Mitsuya et al. (2011) recommended perceptual measures in combination with the altered auditory feedback paradigm to further analyze the phonological mediation of auditory feedback. Vietnamese was chosen for comparison because it differs from English as a tonal language with more vowels. To extend cross-language knowledge, the study involved early bilinguals of English in which the Vietnamese talkers were dominant in English. This is in contrast to Mitsuya et al. (2010, 2011), in which they had late bilinguals of English. Monolingual English talkers are used as controls in the study to prevent confounds from other languages. Due to time limitations and participant fatigue, English vowels /ɪ, æ/ in 'hid, had' and Vietnamese vowel /ɐ/ in 'tăm' were chosen as target vowels for the altered auditory feedback. As well, perceptual measures were conducted for English /ɪ/ only.

Perceptual tasks are required for this cross-language study to examine the phonological mediation of acoustic feedback (Mitsuya et al., 2011). This experiment contains two parts: two perceptual measures and altered auditory feedback production were employed for the various English and Vietnamese vowels. Similar to the Villacorta et al. (2007) study, the perceptual discrimination threshold for changes in F1 was determined. Vowel goodness ratings (similar to Kuhl, 1991) were also obtained for a set of /ɪ/-like vowels where F1 was varied and an estimate of category bounds was determined. Compensation in speech production in response to altered auditory feedback was measured as the magnitude of change in production of F1 during the Hold and various steps in the Ramp phase relative to the Baseline phase.

1.6.1 Hypotheses

The F1 discrimination threshold and vowel goodness may be related to compensation to altered auditory feedback. When the manipulation of auditory feedback exceeds the F1 discrimination threshold and estimated categorical bounds during the Ramp phase, the talkers should start to compensate. Furthermore, compensation should occur in the opposite direction of the perturbation and continue to increase with increasing perturbations. However, the compensation is expected to be smaller than the magnitude of perturbation.

The Vietnamese talkers in the study learned English at a young age and considered to be early bilinguals. According to the CPH (see section 1.3.1), early bilinguals may perform similarly to the English monolinguals because the Vietnamese talkers are English dominant. It could be hypothesized then that the perceptual and compensation results will be similar between the two groups. However, according to the cross-language interference and SLM theories (see section 1.3.3), the interaction of L1 and L2 elements may cause differences between the two groups. In early bilinguals, L1 and L2 phonic elements have a higher probability to deflect away or dissimilate from each other (see section 1.3.3.2). As a result, Vietnamese bilinguals may perceive and produce English elements differently than English monolinguals. Therefore it could alternatively be hypothesized that there would be differences between English monolinguals and Vietnamese bilinguals in the perceptual and altered auditory feedback tasks. Results of the present study should offer some evidence and insight regarding two reasonable hypotheses and this complex system.

The NRV (Polka & Bohn, 2003; see section 1.5.2) may play a role in perception and compensation. During the vowel goodness ratings (see section 2.3.1), the exemplars were presented in random order; as a result, the direction of the shift change was controlled for. The talkers were presented with negative and positive shifts of the English and Vietnamese vowels during the altered auditory feedback. If there was a difference in the negative and positive shifts, the NRV perspective may be able to explain the directional asymmetry. If the shift was towards the peripheral area of the vowel space, it is hypothesized that greater compensation would occur because feedback error discrimination would be easier.

Chapter 2

2 METHODS

2.1 Procedure summary for participants

Upon arrival of the participant, the experiment was explained to her/him and he/she were asked for informed consent and to complete various questionnaires. An audiometric screening was then performed to determine if the participant had normal hearing thresholds. Two perceptual tasks were then performed in a quiet, laboratory environment to determine vowel goodness ratings and F1 discrimination thresholds. The participant was then seated in a sound booth to perform altered auditory feedback tasks. A summary of the experiment and compensation for the participant's time were provided when all tasks were completed. All questionnaires and experimental procedures were approved by the University of Western Ontario Ethics Board (Appendix A). Below are further details of the experiment.

2.2 Participants

Forty-seven participants were recruited from the University of Western Ontario, city of London and the Greater Toronto Area and were divided into two groups. Group one included 21 native Vietnamese speakers with ESL [17 females, 4 males; ages: 18-25 yr, mean: 20.76 yr, standard deviation (SD): 2.12 yr]. All the Vietnamese participants learned ESL in Ontario. The Vietnamese talkers went to English-speaking schools since Kindergarten and as a result they were surrounded by English-speaking individuals and by English social media. Group two included 21 English monolingual talkers (17 females, 4 males; ages: 18-35 yr, mean: 23.33 yr, SD: 3.94 yr). English participants had learned English in Ontario, Saskatchewan, and British Columbia. For each ear, hearing thresholds were measured at octave intervals between 250 Hz and 4 kHz. Individuals generally had normal thresholds (≤ 20 dB HL). However, sound leakage at low frequencies from the supra-aural headphones (TDH-296) may occur, as well as low frequency masking from ventilation noise to the left of the participants in the sound booth. As a result, we included participants who had thresholds up to 30 dB HL at low frequencies (see Table 1). Additionally there was one participant with a slightly elevated threshold at 4000 Hz in one ear. This was not expected to influence the results using supra-threshold speech so the participant was retained. Each participant was tested in a single session. No participants had known language, hearing, or speech impairments. Data from five participants were discarded because they learned English in the Maritimes, were not born in Canada, or had a history of a lisp or a stutter.

2.3 Perceptual tasks

2.3.1 Vowel goodness and category bounds

Goodness is defined as the ability of an exemplar of a specific sound to fit into its respective category (Kuhl, 1991). Vowel goodness ratings were used to determine the goodness of various exemplars of the vowel /ɪ/ in ‘hid’. Eleven versions of “hid” were created, with one unaltered version and ten altered versions. These altered versions were created by filtering similar to that used during online formant shifting, which is described in section 2.4.4. To create the altered version, F1 of the unaltered “hid” was shifted upwards in 20 Hz steps to +200 Hz (i.e. +20, +40, +60, +80 ... +200 Hz) on a continuum towards /ɛ/. The eleven sounds were played in a permuted order, nine different times, providing 99 trials in total. The first four iterations of the set of sounds were discarded to allow the participant to know the full range of /ɪ/-like sounds. The participant was asked to rate each sound of “hid” on a scale of 1 to 7, where 1 is a very poor version and 7 is an excellent version of “hid”. These goodness ratings were used to determine the participant’s perceptual organization of the vowel /ɪ/: the continuum between a good and bad /ɪ/. The vowel /ɪ/ category bound towards /ɛ/ was determined by fitting a sigmoid function to the goodness data and then determining the shift size where the sigmoid achieved half its height.

Table 1: Elevated hearing thresholds of participants.

Numbers of participants are combined for English and Vietnamese participants

Ear	Right			Left		
Frequency (Hz)	250		4000	250		500
Threshold sound level (dB)	25	30	25	25	30	25
Number of participants	4	1	1	17	2	2

2.3.2 F1 discrimination threshold

An F1 discrimination threshold was defined as the minimum change in F1 that the listener detected. A two-alternative forced choice test (2AFC) was used to determine the F1 discrimination threshold for /ɪ/ in ‘hid’, with shifts of F1 in the positive direction towards /ɛ/ in ‘head’. A continuum of “hid” was created by shifting F1 upwards in 5 Hz steps using a method similar to that done online (see section 2.4.4). The unaltered “hid” was produced by a young adult male who spoke English as his first language. An adaptive AXB 2AFC program called Dinosaur created by Dorothy Bishop (n.d.) was selected to perform this task. In Dinosaur, the participant was asked to determine which sound, the first or last, was like the middle sound, the unaltered “hid”. As the Dinosaur program progressed, the discrimination between sounds became more difficult. When the listener made an incorrect choice, the tasks were made easier by having the listener hear a larger shift. When the listener made correct choices, the tasks were made harder by having the listener hear a smaller shift. When Dinosaur was completed, after eight reversals, the listener’s F1 discrimination threshold was found by averaging the shift magnitude for the final four reversals (see Figure 5 for an example of an individual’s performance).

2.4 Altered auditory feedback

2.4.1 Equipment

Participants were prompted to speak when the target word appeared on a computer screen at a rate of approximately once every two seconds (Purcell & Munhall, 2006a). Participants wore a Shure WH20 headset microphone (see Figure 6). The microphone signal was amplified with a microphone amplifier (Tucker-Davis Technologies MA3) with a +20 dB gain switch active and adjustable gain set individually as described below. The signal was low pass filtered with a cut-off frequency of 4500 Hz (Frequency Devices type 901). The analogue signal was then digitized at a 10 kHz sampling rate with 18-bit precision (National Instruments PXI-6289M input/output board). During altered auditory

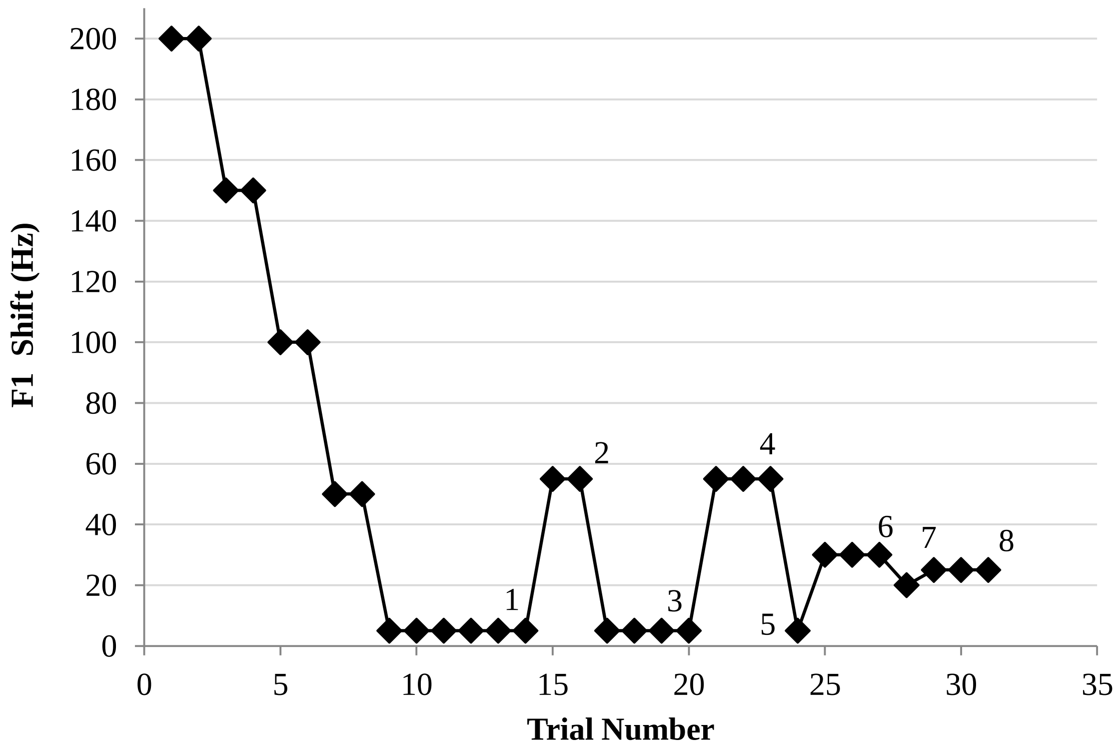


Figure 5: Plot of trials of one participant for the two alternative forced choice test. Numbers 1-8 near the line show where the eight reversals occurred. F1 discrimination threshold was found by averaging the shift magnitudes for the final four reversals. In this case the threshold was 27 Hz.

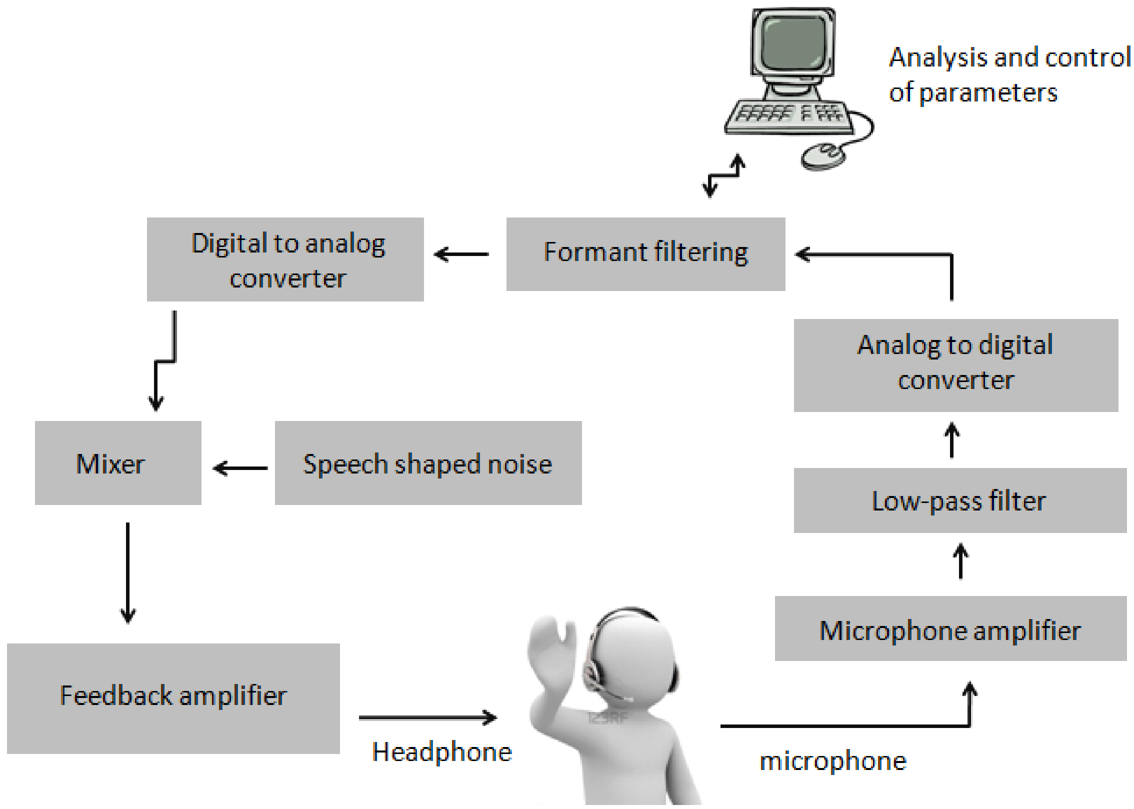


Figure 6: Equipment involved with altered auditory feedback.

feedback, the signal was analyzed and filtered in real time to create the formants shifts (National Instruments PXI-8106). The digital signal was converted back to an analogue sound at 10 kHz with 16-bit precision by the National Instruments PXI-6289M and routed to a Madsen Itera audiometer for amplification. For each participant during practice trials, the microphone MA3 amplifier gain was adjusted between 20 and 40 dB so that vocal sounds reaching the Madsen Itera input VU meter were approximately 0 dB. With this input level achieved on the input VU meter, the voice signals were presented back to the listener at an amplified level of 80 dBA sound pressure level (SPL) using Sennheiser “HD 265 linear” headphones. Background speech shaped noise of 50 dBA SPL was also added by the Madsen Itera audiometer. The purpose of this noise was to hide small imperfections that may have occurred during filtering. All equipment reported was similar to Purcell and Munhall (2006b).

2.4.2 Target vowels

Due to time limitations only a few English and Vietnamese vowels were manipulated. The vowels English /ɪ/, English/æ/, and Vietnamese /ɐ/ were chosen as target vowels for reasons described below. An overlay of the Canadian English and Vietnamese vowel spaces was created to determine which vowels to manipulate (Figure 7). English /ɪ/ was chosen because Purcell et al. (2011) reported an asymmetry in compensation: the positive shift had greater compensation than the negative shift. As well, the compensation in the negative shift was generally smaller than all other negative shift conditions for other vowels. This was interesting because /ɪ/ is not a corner vowel in the Canadian English vowel space and the small compensation was not expected. The current study used a different Ramp phase than Purcell et al. (2011) to further examine this asymmetry in compensation. As well, in the CPS of Vietnamese bilingual talkers, English /ɪ/ is near English /ɛ/ and Vietnamese /ɛ/ where these vowels have higher F1 frequencies (see Figure 8A). There are no vowel categories with lower F1 frequencies in the immediate vowel space area because the F2 of vowels with lower F1 (like /i/) are significantly different from F2 of /ɪ/. The English /ɪ/ category may have assimilated into one of these /ɛ/ categories for the bilinguals. Similarly, English /æ/ was chosen based on the surrounding vowel categories in the CPS of Vietnamese talkers. English /æ/ is

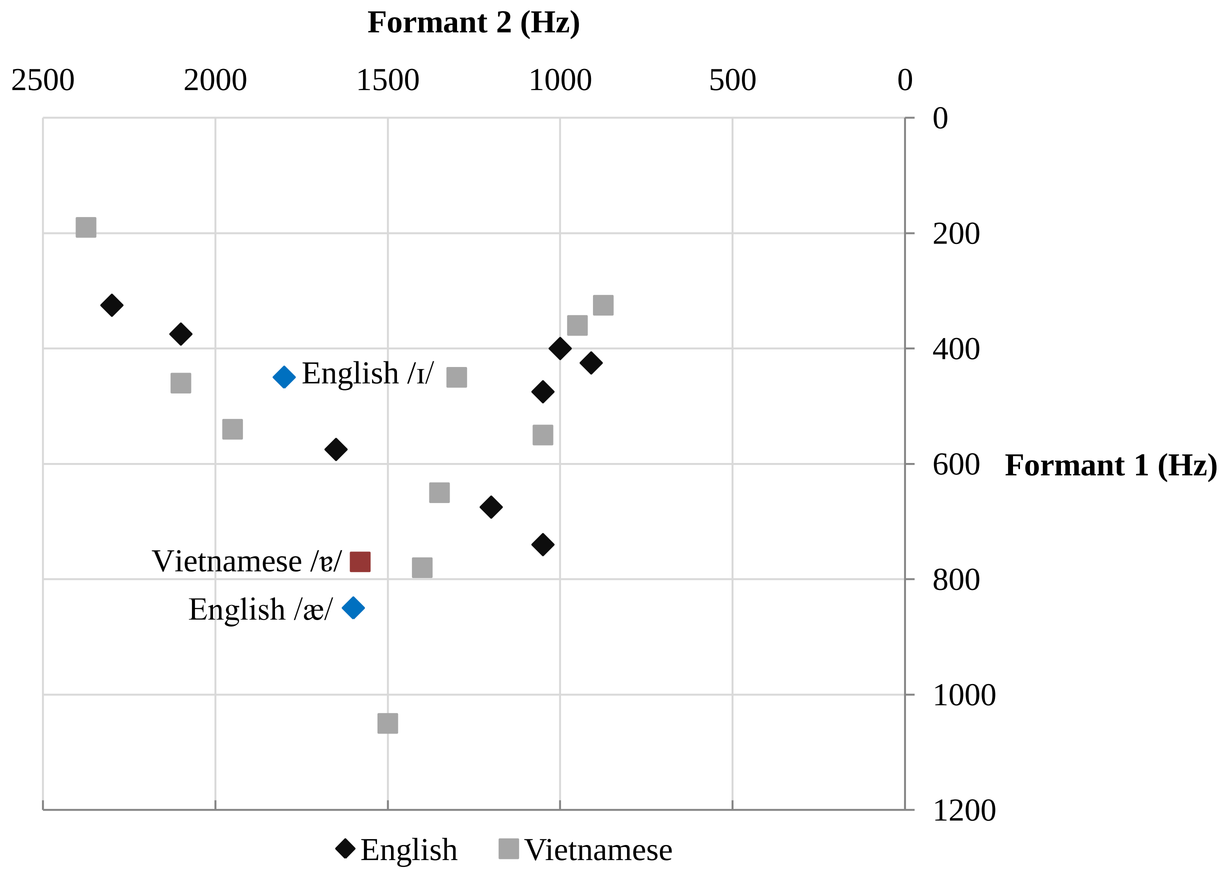


Figure 7: Combined English and Vietnamese vowel space.

Black diamonds represent English vowels. Grey squares represent Vietnamese vowels. English target vowels /ɪ, æ/ are highlighted in blue. Vietnamese target vowel /ɐ/ is highlighted in red. Axes are the first and second formants. Vowel locations are based on the English vowel space by Mendoz-Denton et al. (2001) and the Vietnamese vowel space by Santry (1997).

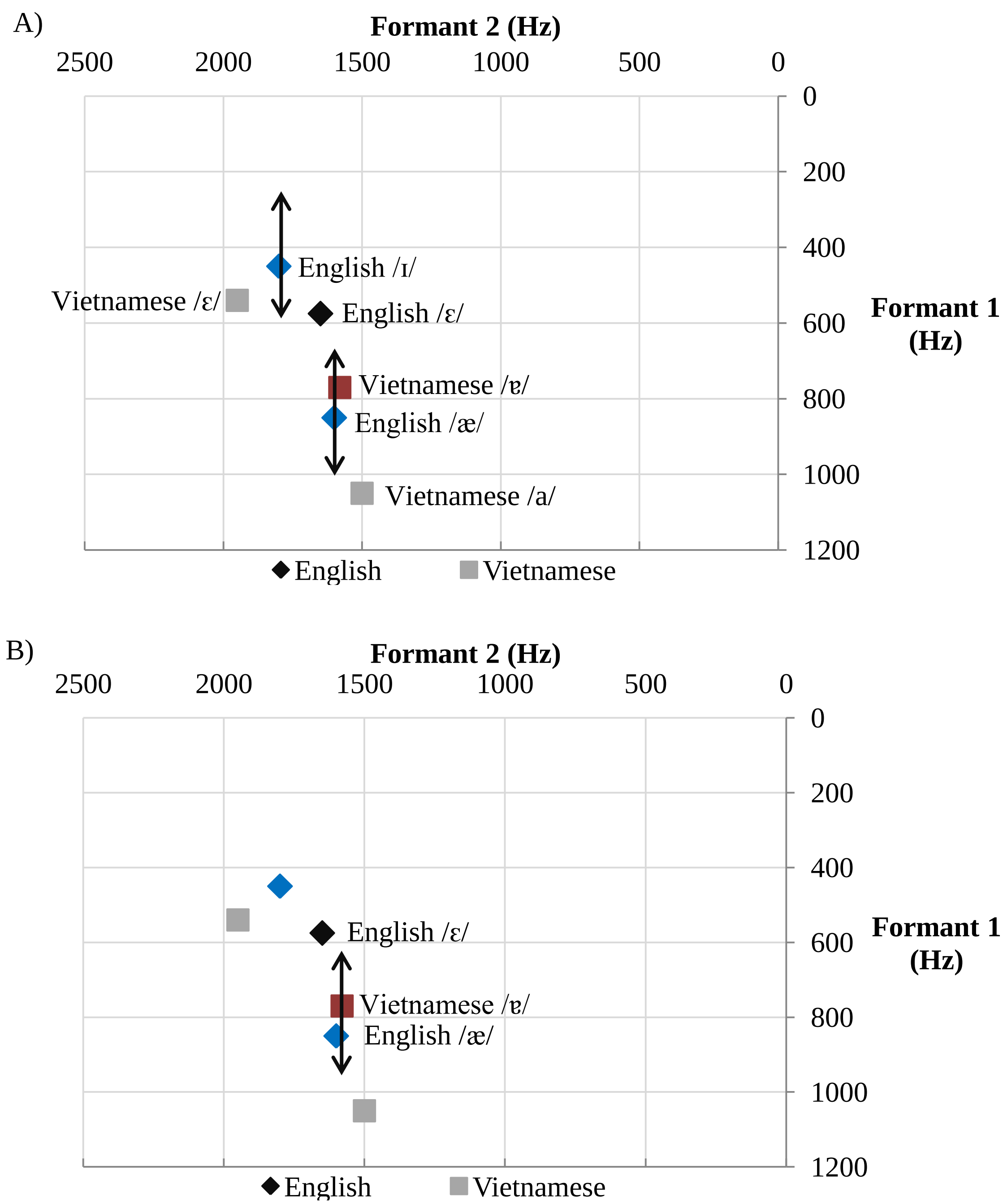


Figure 8: Vowel location of manipulated vowels.

A) English vowel /ɪ/ and /æ/ B) Vietnamese vowel /ɐ/. Black diamonds represent English Vowels. Grey squares represent Vietnamese vowels. English target vowels /ɪ, æ/ are highlighted in blue. Vietnamese target vowel /ɐ/ is highlighted in red. Black arrows indicate changes in F1 that are introduced in auditory feedback. Axes are the first and second formants. Vowel locations are based on the English vowel space by Mendoz-Denton et al. (2001) and the Vietnamese vowel space by Santry (1997).

surrounded by Vietnamese vowels, where Vietnamese /ɐ/ has a lower F1 frequency and Vietnamese /a/ has a higher F1 frequency (See Figure 8A). The English /æ/ may have assimilated into one of the Vietnamese vowel categories. In contrast, in the English monolingual talkers, English /æ/ is a point vowel where there are no vowel categories with higher F1 frequencies. Vietnamese /ɐ/ was also chosen based on the surrounding vowel categories in the CPS of Vietnamese talkers: Vietnamese /ɐ/ is surrounded by English /ɛ/ and English /æ/ (see Figure 8B). Overall, vowels chosen to be used in the experiment were based upon their location in the CPS of Vietnamese talkers and on past research.

2.4.3 Screening procedure and model order estimation

Participants were seated in a sound booth (Eckoustic C-26) in front of a computer monitor. They were asked to speak in their normal voice and to keep their loudness and pitch reasonably stable as they produced each of the prompted words on the screen.

Linear predictive coding (LPC) is a common approach to estimate formants in speech signals (O'Shaugnessy, 1988). The LPC method determines linear filter coefficients which can predict the current speech sample from a weighted combination of previous samples. When the filtering characteristic of these coefficients is represented in the frequency domain as a spectrum, it is like a spectral envelope fitted over the actual speech spectrum (see Figure 9 for an example). The peaks in this LPC envelope give the formant estimates, where the number of formants is set by the model order. Before the altered auditory feedback, a screening procedure was used to determine the best model order (BMO) for producing stable formant estimates. Six tokens of the target vowel from the screening period were evaluated with LPC model orders 8 to 12 and the order that produced the least variance in estimates of F1 and F2 was selected, as described in the next paragraph. Determining the BMO reduced the errors in tracking F1 throughout the altered auditory feedback.

The model order with the lowest F1 and F2 standard deviations was chosen as the BMO for each target vowel during the screening procedure. For example, Figure 10 depicts the

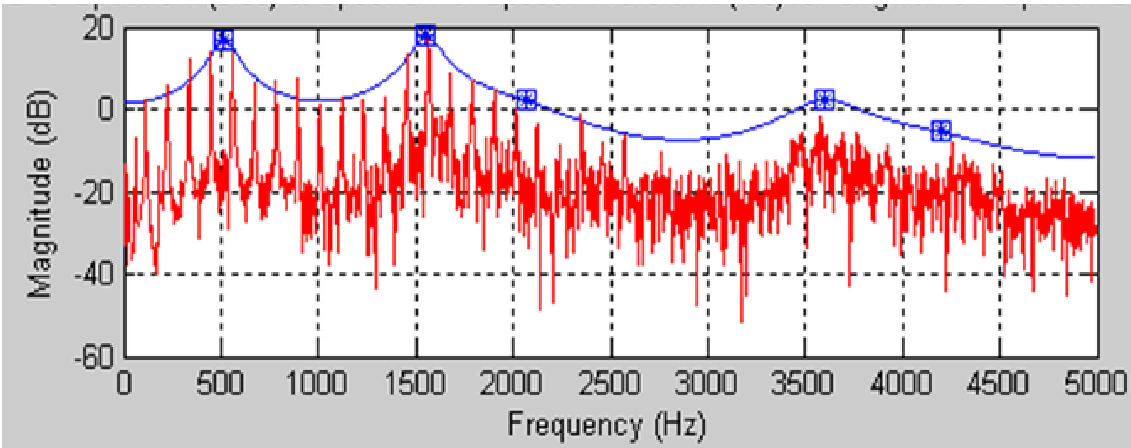


Figure 9: Formant estimates.

Speech signal is represented in red (mostly voice harmonics for frequencies < 2500 Hz) and a LPC spectral envelope is represented in blue. The first two blue squares at approximately 500 and 1500 Hz are the estimates of F1 and F2, respectively.

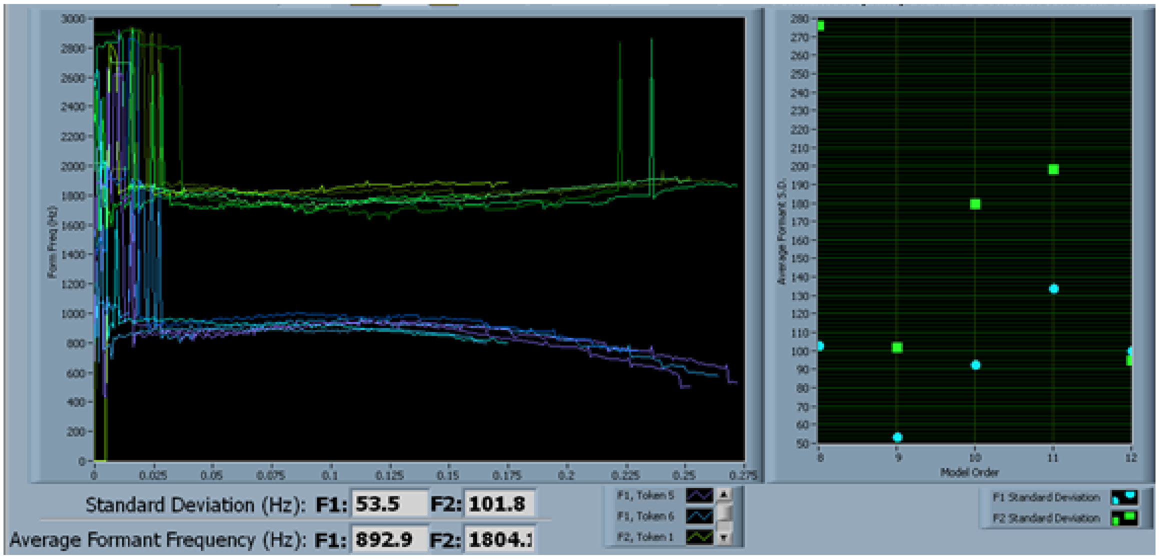


Figure 10: Formant estimates and standard deviations for English /æ/ in 'had' for one participant.

F1 is shown in blue and F2 is shown in green. The left graph represents the F1 and F2 tracks over time (seconds) for the vowels excised from six utterances for the currently selected model order 9. The right graph gives F1 and F2 standard deviations for all model orders. The criteria for best model order was having the most stable formant tracks which was defined as that with the lowest F1 and F2 standard deviations (model order 9 in this example).

model orders evaluated for a participant producing the target English vowel /æ/. The formant tracks derived from six utterances using model order 9 are on the left side of Figure 10. The vowels were excised from these six utterances and LPC produced the formant tracks shown to last approximately 275 ms in this example. Formant standard deviations were calculated for each utterance and then averaged for each model order. These standard deviations are shown on the right side of Figure 10 for all model orders. For this participant, model order 9 produced formant estimates with the smallest standard deviations in F1 and F2. This BMO would be used to estimate formants for all utterances of this specific vowel during the main altered auditory feedback experiment.

The screening process involved the participants repeating seven English monophthongs /i, ɪ, e, ɛ, æ, ɑ, u/ in an /hVd/ context. The set of words were permuted and the complete set was presented randomly, six times sequentially; therefore, there were 42 prompted words in total. A BMO was selected for each of the desired target vowels: /ɪ/ and /æ/.

The Vietnamese talkers also produced eight Vietnamese monophthongs /i, e, ɛ, ɐ, a, ɤ, ʌ, u/ in a /CV/ or /CVC/ context. There were differences in /CV/ or /CVC/ context for the Vietnamese vowels because the Vietnamese participants had difficulty in the pronunciations of the words in isolation when the context was the same. The set of words were permuted and the complete set was presented randomly, six times sequentially; therefore, there were 48 prompted words in total. A BMO was selected for the Vietnamese vowel /ɐ/.

2.4.4 Online voice detection and formant shifting

Altered auditory feedback was achieved by filtering the voice in real-time. Filtering was applied during the voiced part of each utterance. To detect the onset of voicing in each trial, a statistical amplitude threshold technique was used. This technique involved determining a mean and standard deviation microphone input level from a quiet period prior to the prompt. Voice onset was determined when the microphone input level exceeded the mean microphone input level from the quiet period by six standard deviations. Once voicing was detected, the applied filter used coefficients determined from real-time LPC formant estimates (Figure 11A), which were updated every 900 μs. The speech signal was processed through two filters simultaneously to create the formant manipulations (Figure 11B). One filter de-emphasized the voice harmonics near the produced F1, while the other emphasized voice harmonics at the desired new F1 location. The result of these filters created an altered vowel sound where F1 was shifted in frequency (Figure 11C). Formant calculations included past samples for an effective delay in formant estimates and corresponding filter coefficients of 10 to 20 ms. Voice detection and formant shifting were performed as described by Purcell and Munhall (2006b).

2.4.5 Offline formant analysis

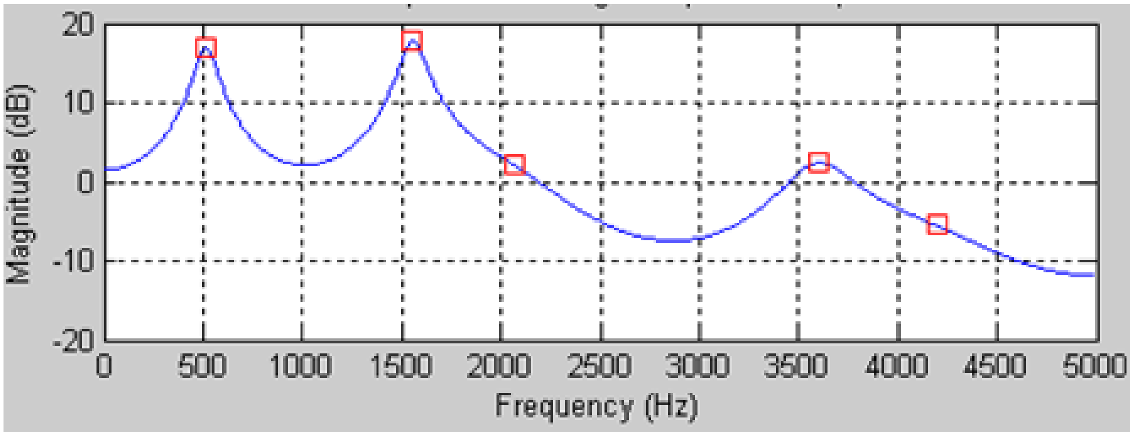
Trials where the prompt was read incorrectly were removed from the analysis. A semi-automated process was used to trim every utterance before and after the vowel (Figure 12). The vowel boundaries were examined by the experimenter and corrected if required.

Single steady-state values of F1, F2 and third formant (F3) for each utterance were estimated offline. This single value for each formant was calculated by averaging the estimates from 20% to 80% of the way through the utterance for that formant. The first and last 20% of the vowel are not used in the analysis because vowel transitions have unstable formants. Occasionally, formants were incorrectly categorized as another by the estimation algorithm (i.e. F1 being categorized as F2, etc.). These errors were found and corrected by examining a graph (Figure 13) with all the F1, F2, and F3 values for each participant. As seen in Figure 13, formant values were displayed in the order that they were produced during the experiment. With this approach, outliers could be readily identified and relabeled. Offline formant analysis was similar to Munhall et al. (2009).

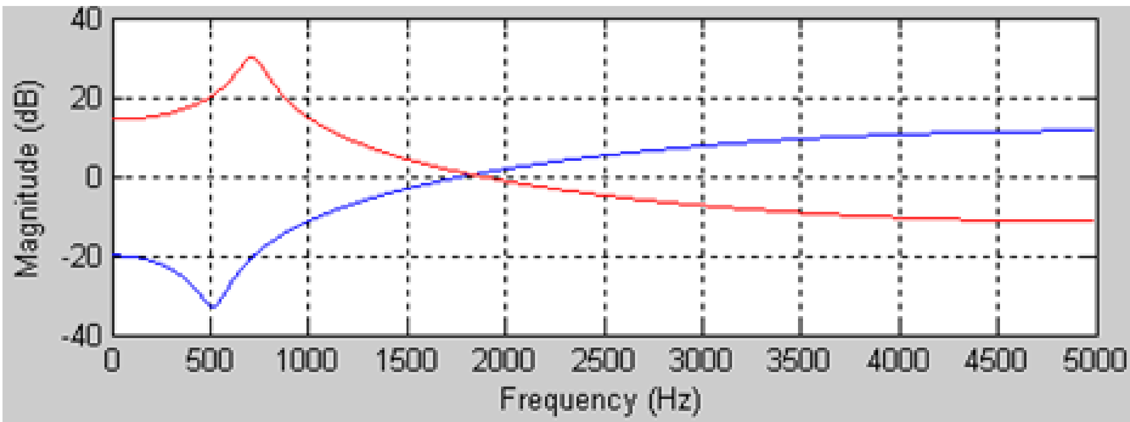
2.4.6 Procedure and experimental conditions

Before entering the sound booth, informed consent was obtained and participants completed three questionnaires (medical history background, language and music history;

A)



B)



C)

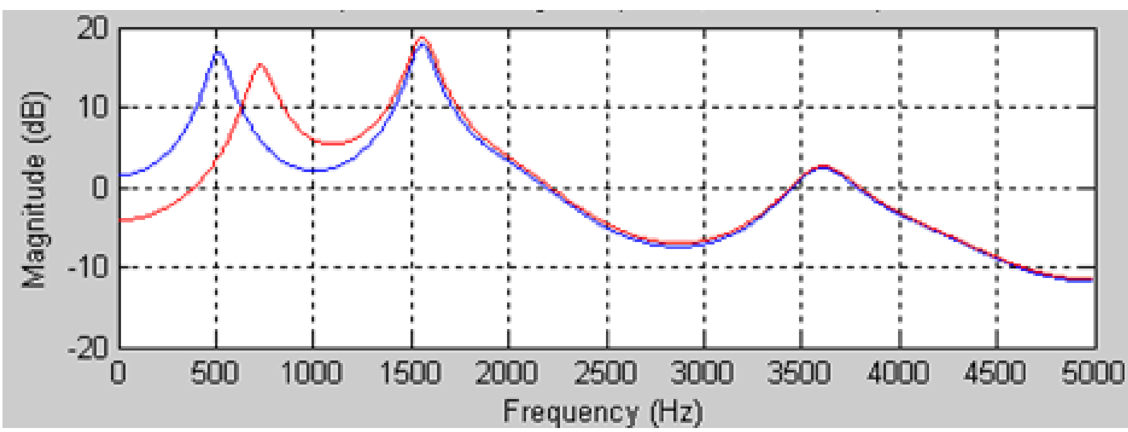


Figure 11: Formant manipulations.

A) LPC spectrum representing the original spectrum of the speech sample with the best model order. Red boxes show the formants with F1 at 500 Hz. B) Red line represents a filter which emphasizes existing voice harmonics at the new F1 (500+200=700 Hz). Blue line represents a filter which de-emphasizes voice harmonics in the speech sample at the produced F1. C) Red line presents the LPC spectrum of the new vowel where F1 is changed by +200 Hz. Blue line represents the LPC spectrum of the original speech.

A)

B)

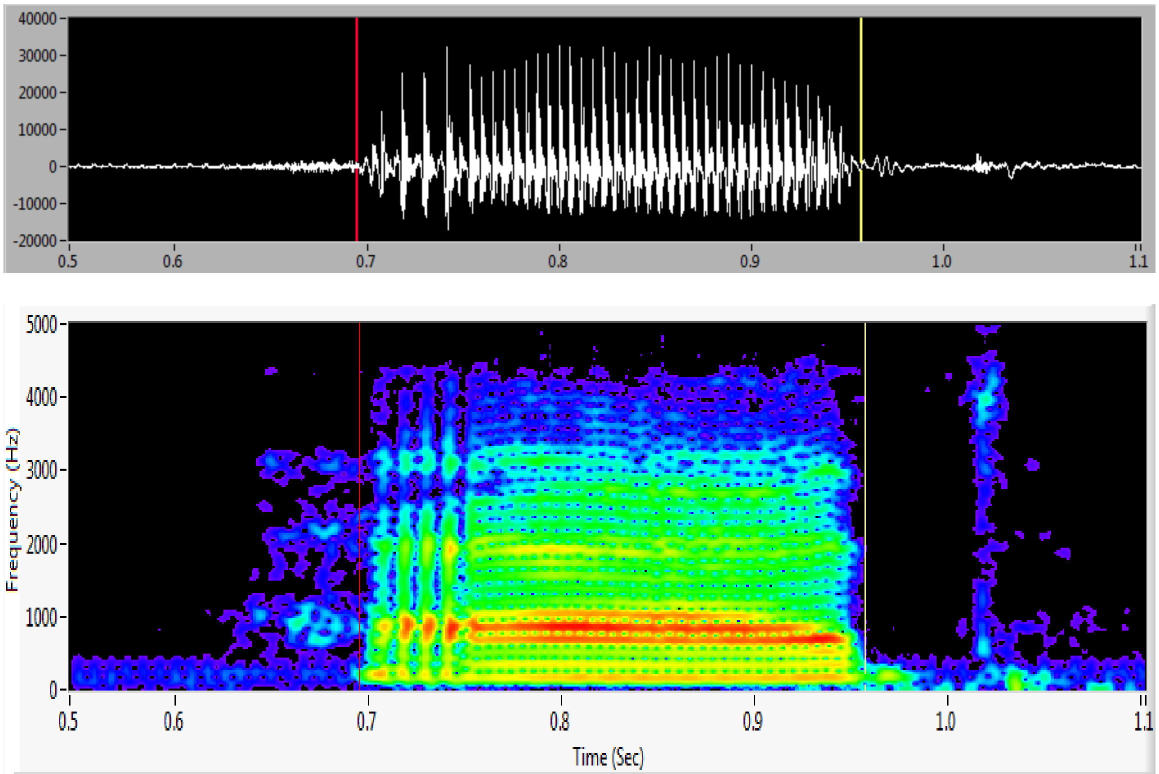


Figure 12: Vowel boundaries for English /æ/ in ‘had’ for one utterance.

A) Time Waveform. B) Spectogram. Red and yellow vertical lines represent the beginning and end of the vowel, respectively**.**

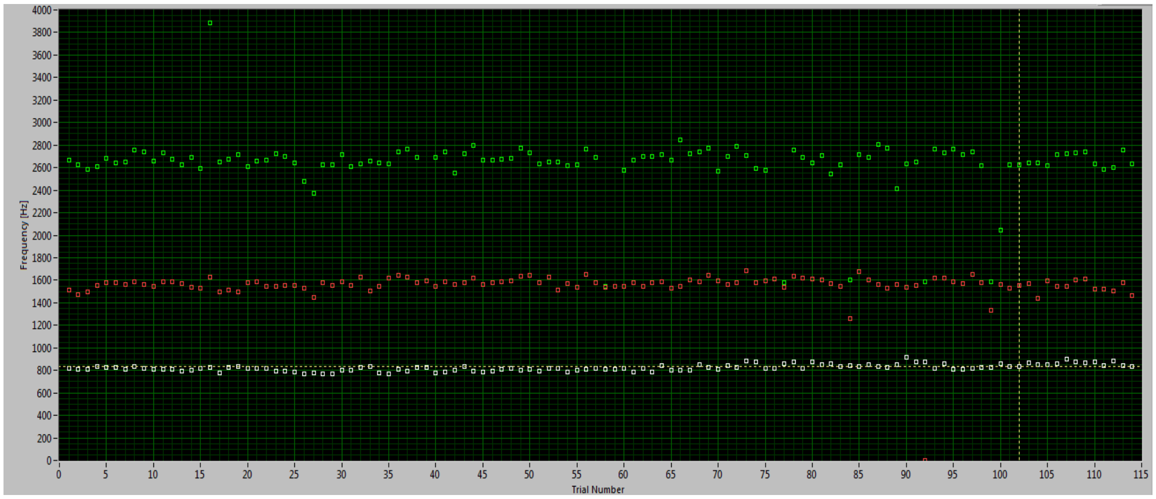


Figure 13: Formants across trials for one participant for English /æ/ in 'had'.
Graph represents F1 (white), F2 (red) and F3 (green) across all trials for English /æ/ in 'had' during the negative shift. If large formant errors are present, corrections are made in this graph by relabeling or removing individual formant estimates. The horizontal axis shows the trial number and the vertical axis shows formant frequency.

see Appendix B). Medical history background was completed because the study required healthy, normal adults. Language history was completed because the study was limited to English monolingual and Vietnamese bilingual talkers, with a minimum grade 12 high school education in an English speaking school in Ontario and Western Canada. Music history was collected from the participants because if there was an anomaly in the data, music history may have provided a reason. A pure-tone audiometric screening was then conducted to determine if the participant had normal hearing thresholds (≤ 20 dB HL) at octave intervals between 250 Hz and 4 kHz using TDH-296 headphones and a Madsen Itera Audiometer. Participants were fitted with noise attenuating headphones (Sennheiser HD265) and sat in front of a notebook computer to perform the vowel goodness ratings and Dinosaur test in a quiet laboratory environment.

In the sound booth, talkers were seated in a comfortable chair and provided with bottled water for consumption between experimental conditions if required. The English vowel space was first collected. Then, vowels for the targeted vowel condition were collected to determine BMO and amplification adjustments. The order of the vowels that were presented was English /ɪ/, English /æ/, and if the participant was Vietnamese, /ɐ/ was last. English /ɪ/ was presented first because the perceptual tests manipulated English /ɪ/. If the participant decided to leave the experiment early, speech perception and compensation data from English /ɪ/ was most likely to be collected. In the English /ɪ/ condition, the F1 positive shift was always presented first, followed by the F1 negative shift. Shifts were counterbalanced for the English /æ/ and Vietnamese /ɐ/. Talkers ran through the altered auditory feedback experiment twice for all vowels with F1 shifted positively and negatively. In each directional shift, there were four phases where the talkers said the word “hid”, “had” or “tăm” 114 times, as shown in Figure 14. In the first phase, Acclimatization (first 15 utterances), participants received normal feedback. These utterances were discarded during analyses. In the second phase, Baseline (utterances 16-35), participants received normal feedback. In the third phase, Ramp (utterances 36-95), F1 was perturbed with the magnitude of the perturbation increasing by 20 Hz after every sixth utterance, resulting in a 200 Hz shift by utterance 90. Finally, in the Hold phase (utterance 96-114), the F1 perturbation at 200 Hz was held constant. English passages were read by the participants with their headphones off to normalize their speech

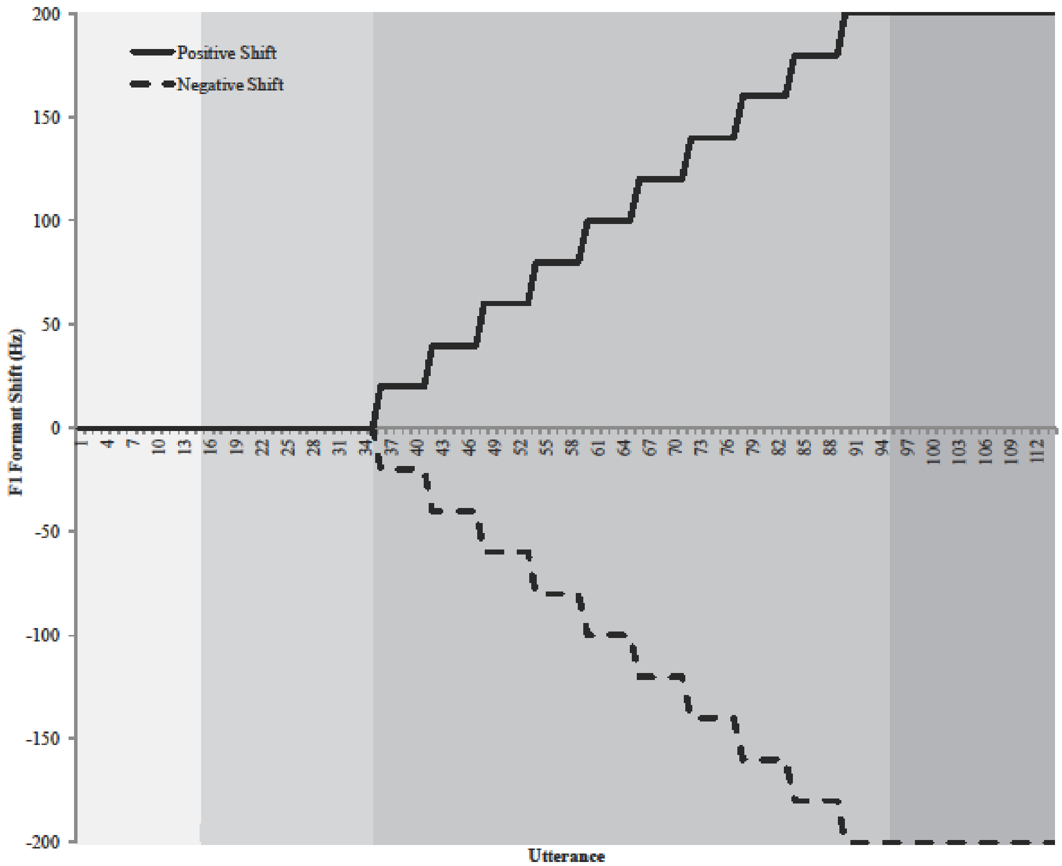


Figure 14: Schematic procedure of F1 formant shifting.

The solid line represents the experimental condition where F1 of the test vowel was shifted upwards. The dash line represents the experimental condition where F1 of the test vowel was shifted downwards. The four phases of the experiment (Acclimatization, Baseline, Ramp, and Hold) are represented with increasing darkness in the shading.

productions after each directional shift, as well as before and after vowel space collections. Vietnamese participants were presented with Vietnamese passages during the Vietnamese part of the study. Common reading passages that are used in the literature (Watt, 2006) and the text passage from the Occupational-Therapy Adult Perceptual Screening Test (OT-APST; Cooke, n.d.) were chosen (Appendix C). Direct translations of the North Wind and the Sun (Bradlow, 2010) and OT-APST passages were used for Vietnamese talkers (Appendix D). OT-APST passages were chosen because Vietnamese and English versions were available. When the procedure was completed, participants were compensated $5 for every half hour for their time.

Chapter 3

3 Results

Vowel space comparisons were performed to determine vowel space density and vowel category similarities between English and Vietnamese vowels. An analysis of perceptual measures: F1 discrimination threshold, vowel goodness ratings, and vowel category bounds were compared between Vietnamese and English talkers. Following, speech compensation during the phases of altered auditory feedback was determined with group comparisons. Correlations between vowel space density and speech compensation are presented. As well, perceptual measures and speech compensations were compared to try to determine the relationship between speech perception and production.

3.1 Vowel space comparisons

English and Vietnamese vowel spaces were collected from all participants. The F1 vowel distance between a target vowel and its nearest neighbour was determined as a measure of vowel space density. Target vowels were the vowels used in the altered auditory feedback trials: for English /ɪ, æ/ and Vietnamese /ɐ/. The F1 manipulation was of magnitude 200 Hz regardless of actual vowel space measurements for all target vowels during altered auditory feedback. Vowel distances in the negative and positive directions were calculated for each female participant, and then a group average was determined. Vowel productions from males and females were not normalized, as a result, males and females were not grouped into one category for vowel space comparisons. Vowel space distances and comparisons were calculated with English and Vietnamese female talkers only because the sample sizes were large (n=17 for each language group). Separate calculations for vowel space distances for male talkers were not performed because of the small sample sizes (n=4 for each language group). Comparisons between the negative and positive shifts and between Vietnamese and English talkers were performed within the English or Vietnamese vowel space. Comparisons between English and Vietnamese vowel spaces were performed to determine similarities and differences of vowel sounds between the two languages. In the vowel space figures, the center of each ellipse represents the mean F1 and F2 frequencies for that vowel, while the solid and dashed ellipses represent one and two standard deviations, respectively.

3.1.1 Vietnamese vowel space

Vietnamese speakers' F1/F2 values of the Vietnamese vowels are plotted in Figure 15A (female talkers) and Figure 15B (male talkers).

During the negative shift, /ɐ/ in 'tăm' was being shifted on a continuum towards /ɤ/ in 'sơn'; the average difference was 165.85 Hz. During the positive shift, /ɐ/ was shifted on a continuum towards /a/ in 'ta'; the average distance was 97.92 Hz between the two vowels. A paired t-test found that the negative and positive distances were not significantly different [$t(16) = 1.38$, $p = 0.187$].

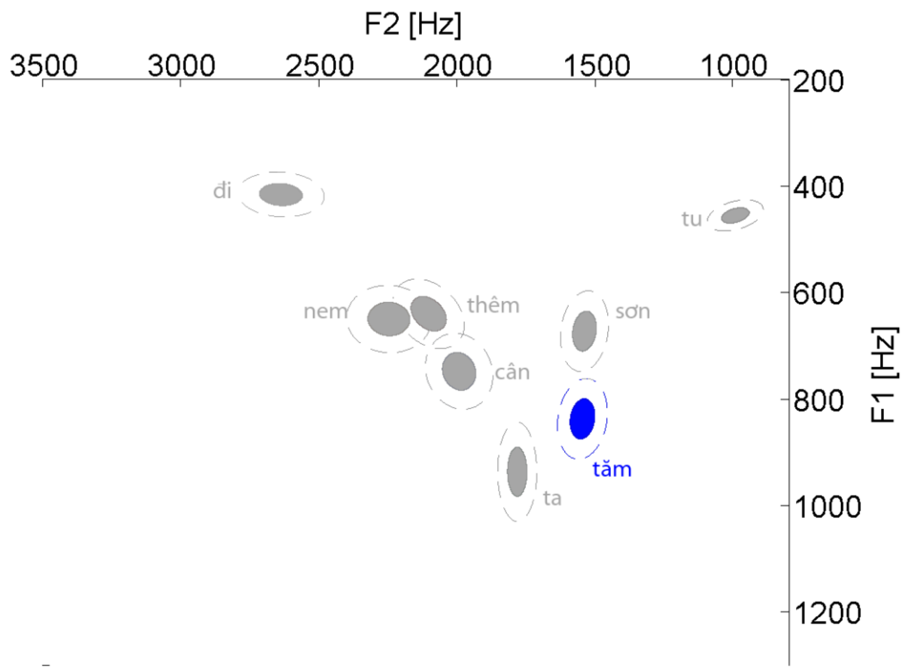
3.1.2 English vowel space

The English vowels produced by the English talkers are plotted in Figure 16A (female talkers) and 18A (male talkers).The English vowels produced by the Vietnamese talkers are plotted in Figure 16B (female talkers) and 18B (male talkers). The English vowel spaces for both groups of talkers were combined into a single vowel space plot in Figure 17 (female talkers) and Figure 19 (male talkers).

F1/F2 productions of English vowels were compared between Vietnamese and English talkers. The English vowel spaces of Vietnamese and English talkers were very similar. Two one-way ANOVA analyses were performed for F1 and F2 to compare the F1 or F2 values of English vowels between English and Vietnamese talkers. Results found /u/ in 'who'd' was significantly different [$F(1, 32) = 8.446$, $p < 0.007$], where F1 of /u/ in Vietnamese talkers were higher than English talkers. All other F1 and F2 values were not significantly different between the two groups.

Similar to the Vietnamese vowel space data, F1 distances between the English target vowels and their nearest neighbours were calculated. In the /ɪ/ in 'hid' condition, the

A)



B)

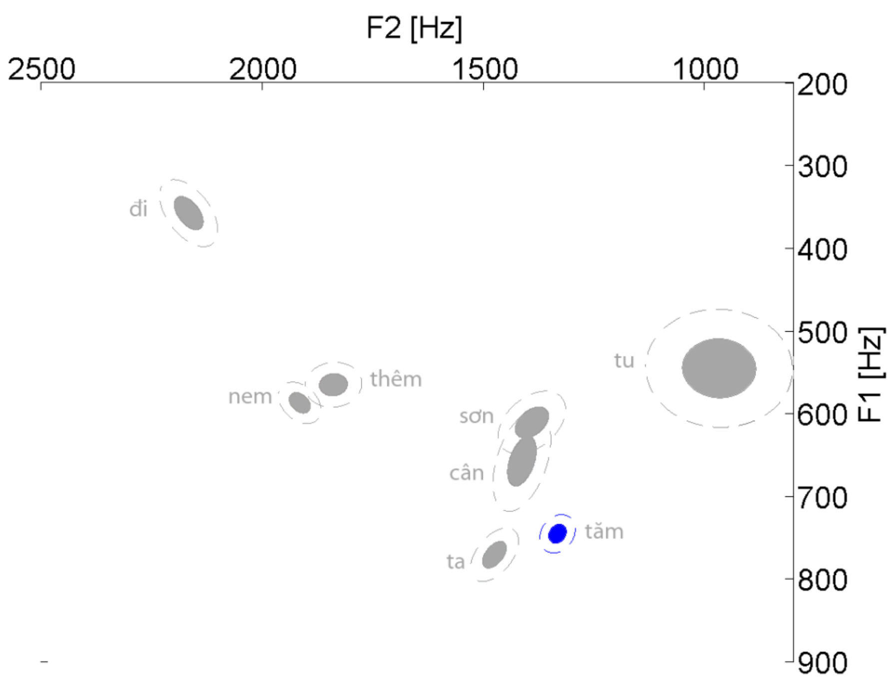


Figure 15: Vietnamese vowel space in a /CV/ or /CVC/ context.

A) Female talkers (n=17). B) Male talkers (n=4). The manipulated vowel /ɐ/ in ‘tăm’ is plotted in blue. The center of each ellipse represents the mean F1 and F2 frequencies. The solid and dashed ellipses represent one and two standard deviations, respectively.

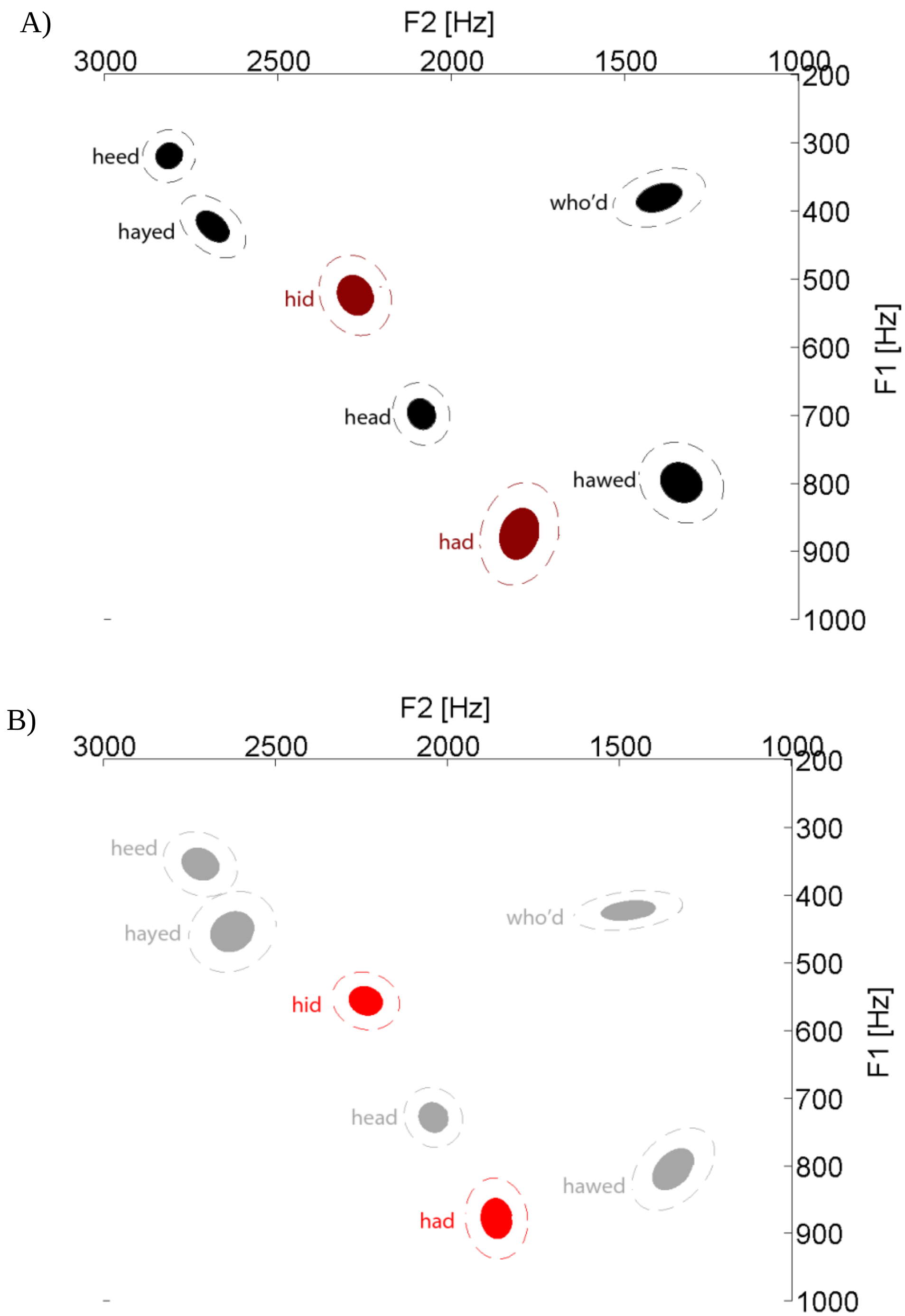


Figure 16: English vowel spaces in an /hVd/ context for female talkers.

A) Female English talkers (n=17). B) Female Vietnamese talkers (n=17). Highlighted in red are the manipulated vowels /ɪ/ in 'hid' and /æ/ in 'had'. The center of each ellipse represents the mean F1 and F2 frequencies. The solid and dashed ellipses represent one and two standard deviations, respectively.

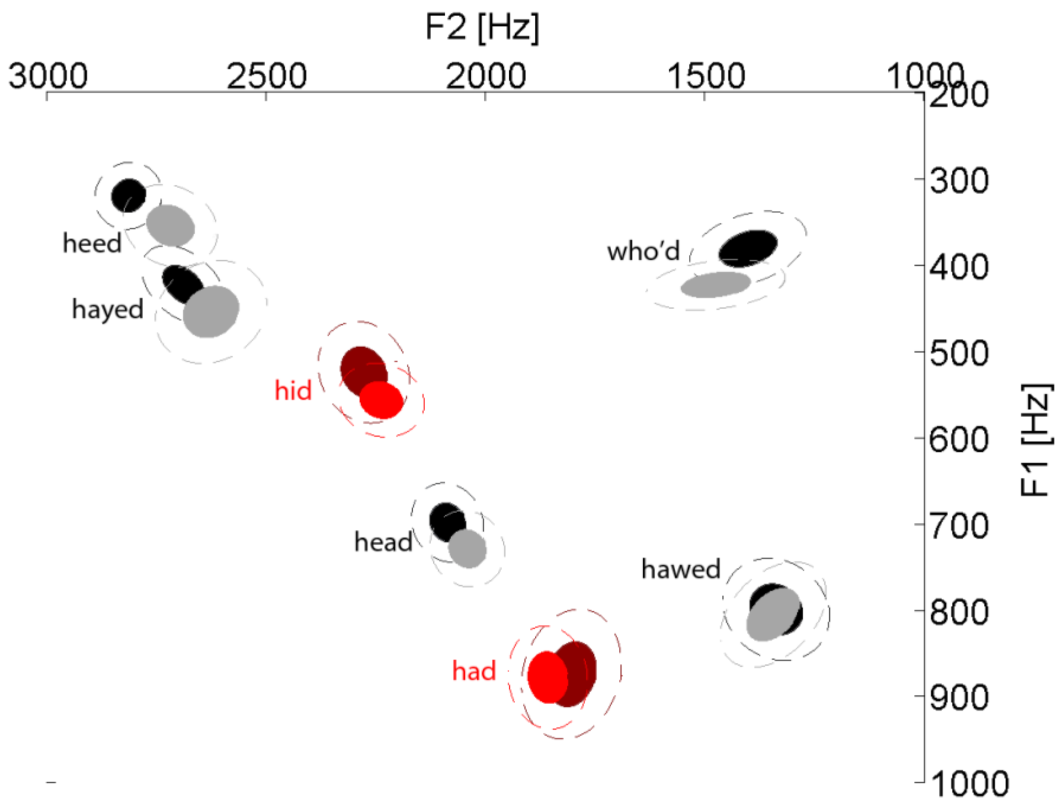


Figure 17: Combined English vowel spaces in an /hVd/ context for English and Vietnamese female talkers.

Black and dark red represent English talkers. Grey and light red represent Vietnamese talkers. Highlighted in the reds are the manipulated vowels /ɪ/ in 'hid' and /æ/ in 'had'. The center of each ellipse represents the mean F1 and F2 frequencies. The solid and dashed ellipses represent one and two standard deviations, respectively.

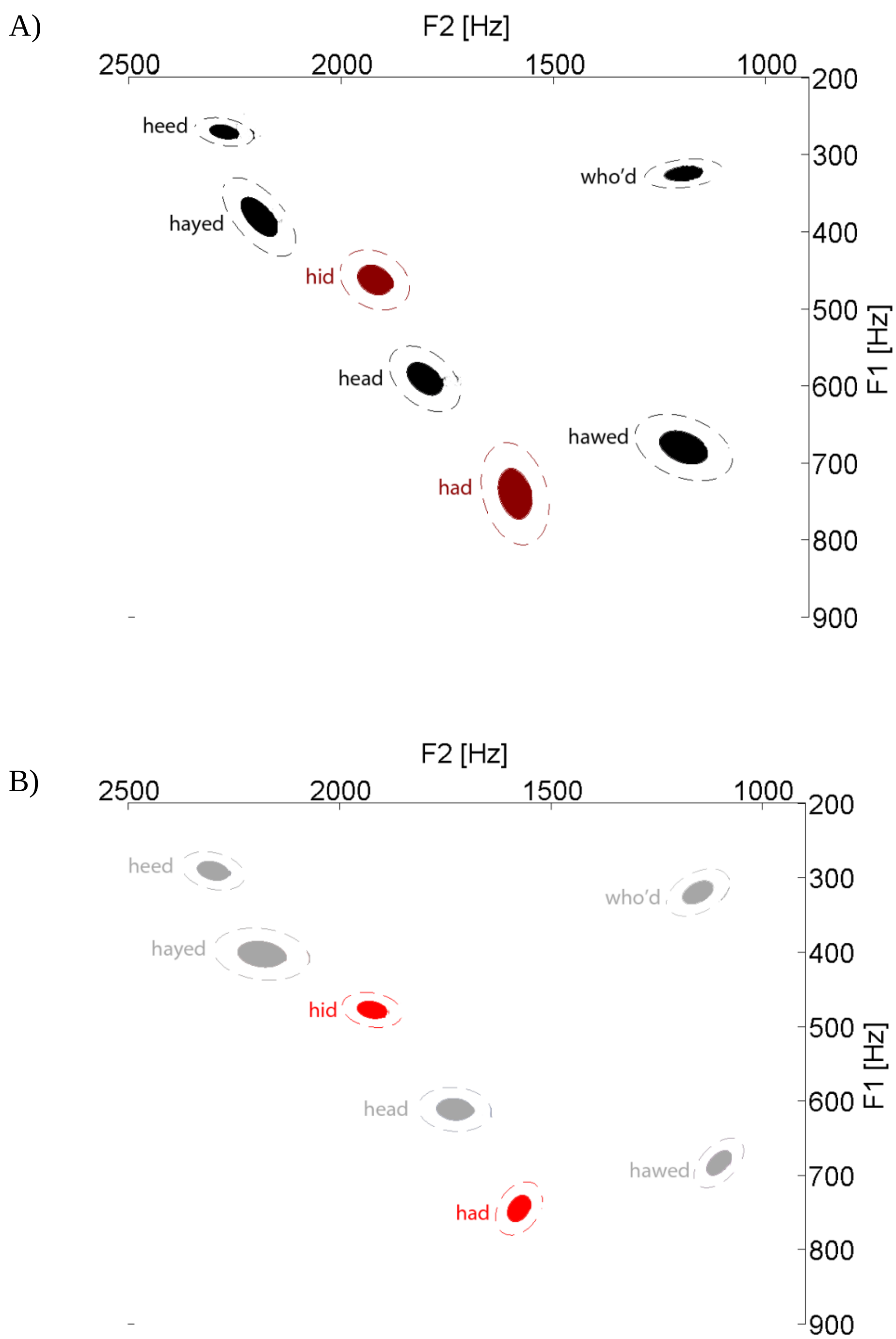


Figure 18: English vowel spaces in an /hVd/ context for male talkers.

A) English talkers (n=4). B) Vietnamese talkers (n=4). Highlighted in red are the manipulated vowels /ɪ/ in 'hid' and /æ/ in 'had'. The center of each ellipse represents the mean F1 and F2 frequencies. The solid and dashed ellipses represent one and two standard deviations, respectively.

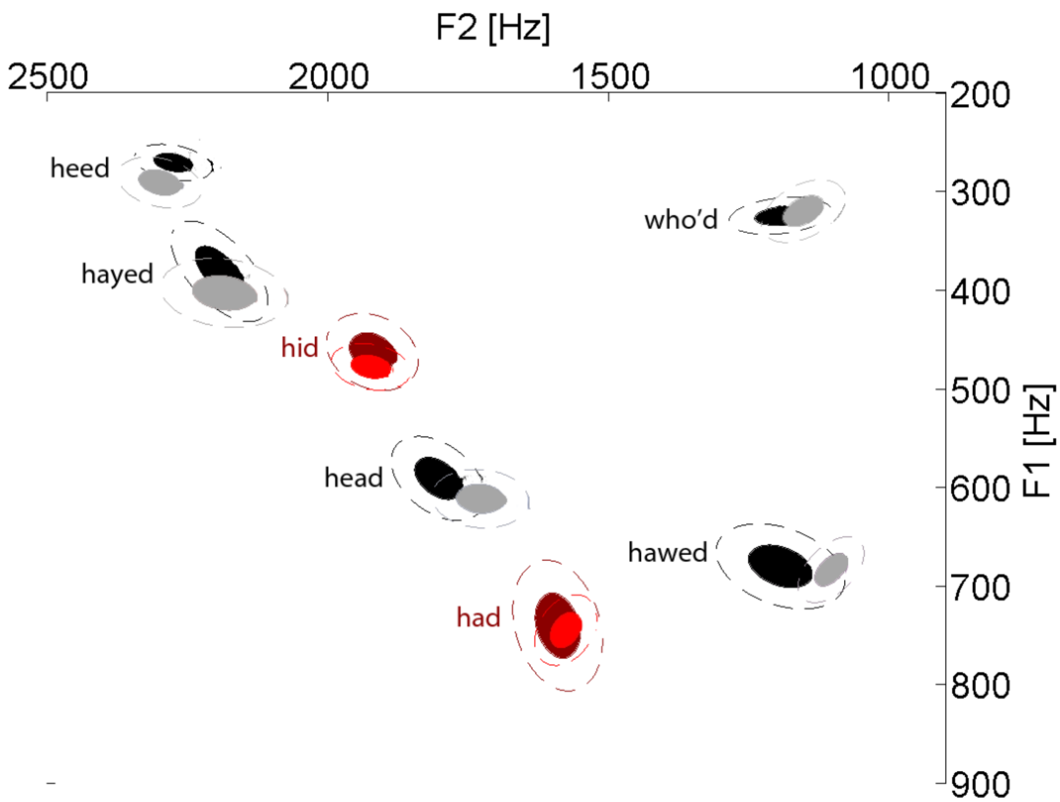


Figure 19: Combined English vowel spaces in an /hVd/ context for English and Vietnamese male talkers.

Black and dark red represent English talkers. Grey and light red represent Vietnamese talkers. Highlighted in the reds are the manipulated vowels /ɪ/ in 'hid' and /æ/ in 'had'. The center of each ellipse represents the mean F1 and F2 frequencies. The solid and dashed ellipses represent one and two standard deviations, respectively.

closest vowel category was /i/ in ‘heed’ in the negative direction. The negative distance was 203.74 Hz for English talkers and 198.84 Hz for Vietnamese talkers. During the positive shift, the closest vowel category was /ɛ/ in ‘head’. The positive distance was 174.07 Hz for English talkers and 175.38 Hz for Vietnamese talkers. An independent samples t-test found that the positive and negative distances were not different between groups [negative: $t(32) = 0.199$, $p = 0.844$; positive: $t(25.995) = -0.097$, $p = 0.923$]. Paired t-tests were performed to analyze differences between the negative and positive distances within groups. There was a significant difference between negative and positive distances for English talkers but not for Vietnamese talkers [English: $t(16) = 2.166$, $p < 0.046$; Vietnamese: $t(16) = 0.741$, $p = 0.469$].

In the /æ/ in ‘had’ condition, /æ/ was shifted on a continuum towards /ɛ/ in the negative direction. The negative distance between /æ/ and /ɛ/ was 174.80 Hz for English talkers and 147.22 Hz for Vietnamese talkers. An independent samples t-test found that the negative distances between groups were not different [$t(32) = 1.375$, $p = 0.179$]. In the common phonological space for Vietnamese talkers (Figure 20), Vietnamese /ʌ/ in ‘cân’ is close to English /ɛ/. A negative shift for English /æ/ may have approached the Vietnamese /ʌ/ category in Vietnamese talkers. The distance between English /æ/ and Vietnamese /ʌ/ was 132.61 Hz in Vietnamese talkers. An independent samples t-test found that the distance from English /æ/ towards English /ɛ/ and towards Vietnamese /ʌ/ was not different [$t(32) = 0.453$, $p = 0.653$]. During the positive shift, /æ/ was shifted by +200 Hz. For the English talkers, /æ/ is a point vowel in their vowel space; as a result, a shift of +200 Hz would not approach another vowel category (see Figure 16). For the Vietnamese talkers, /æ/ was shifted towards Vietnamese /a/; this distance was 59.67 Hz (Figure 20). A paired t-test found the negative distance to the nearest neighbouring vowel was greater than the positive distance [$t(16) = 6.549$, $p < 0.001$] in Vietnamese talkers.

3.1.3 Common phonological space (CPS)

In the CPS for Vietnamese talkers (Figure 20), interaction of English and Vietnamese vowels may occur. Comparisons of F1 and F2 of English and Vietnamese vowels, where they overlapped in the CPS (Figure 20) were completed to determine if an English and a

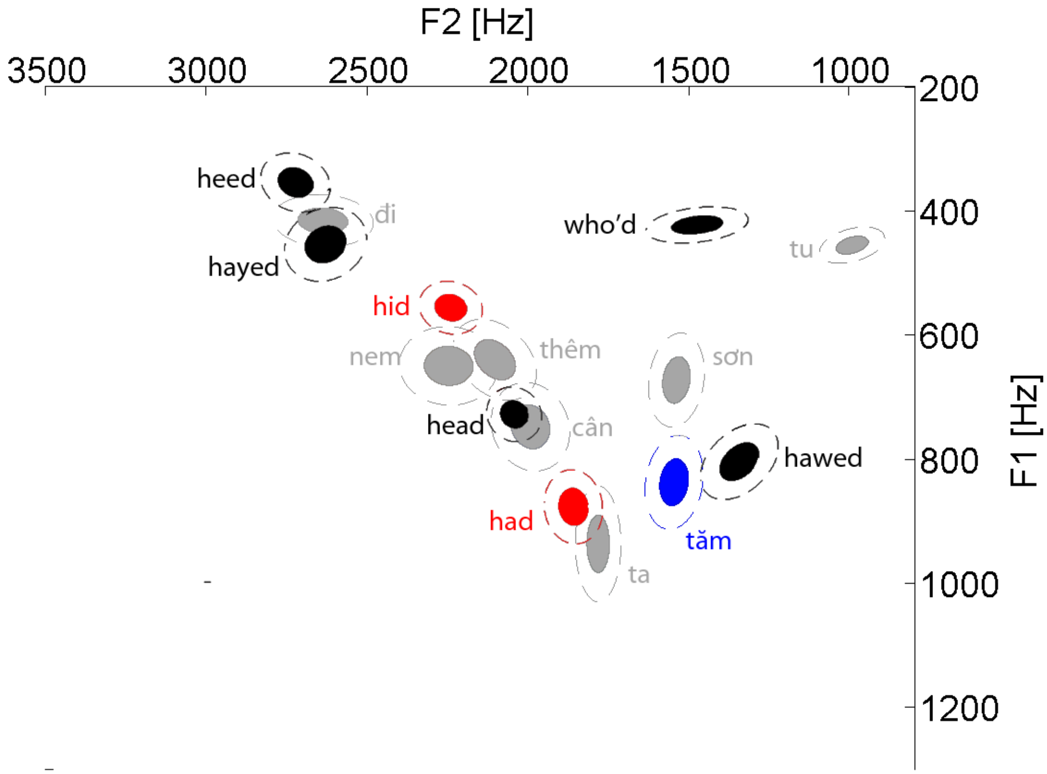


Figure 20: Common phonological space for female Vietnamese talkers (n=17). Highlighted red vowels are the manipulated English vowels /ɪ/in 'hid' and /æ/ in 'had' and blue vowel is the manipulated Vietnamese vowel /ɐ/ in 'tăm'. The center of each ellipse represents the mean F1 and F2 frequencies. The solid and dashed ellipses represent one and two standard deviations, respectively.

Vietnamese vowel category were similar. PCA (Flege, 1995) may have occurred if an English vowel and a Vietnamese vowel had similar F1 and F2 values (Figure 21). Paired t-tests were used to determine if F1 and F2 values of the English and Vietnamese vowel categories were similar. F1 and F2 values for Vietnamese /i/ as 'đi' and English /e/ in 'hayed' were not different [F1: $t(16) = -1.746$, $p = 0.100$; F2: $t(16) = 0.221$, $p = 0.828$]. F1 and F2 values for Vietnamese /ʌ/ in 'cân' and English /ɛ/ in 'head' were similar [F1: $t(16) = -0.549$, $p = 0.591$; F2: $t(16) = 0.324$, $p = 0.750$].

3.1 Speech perception results

3.1.1 F1 Discrimination threshold

The average F1 discrimination thresholds for English and Vietnamese talkers were 28.33 Hz (SD: 5.00 Hz) and 32.90 Hz (SD: 11.72 Hz), respectively. Using an independent samples t-test these thresholds did not differ statistically [$t(40) = -1.644$, $p = 0.108$].

3.1.2 Vowel goodness ratings and category bounds

Group average vowel goodness ratings (Figure 22) indicated that the Vietnamese group gave lower ratings than the English group [Independent samples t-test: $t(20) = 4.922$, $p < 0.038$]. Overall for both groups, higher vowel goodness ratings were given to sounds with small F1 changes, whereas, lower vowel goodness ratings were given to sounds with large F1 changes. The vowel /ɪ/ category bound towards /ɛ/ was determined by fitting a sigmoid function to the goodness data and then determining the shift size where the sigmoid achieved half its height for each participant. The group average category bounds for /ɪ/ on a continuum towards /ɛ/ were determined to be 83.57 Hz (SD: 11.13 Hz) for English talkers and 81.88 Hz (SD: 24.50 Hz) for Vietnamese talkers (Figure 22). An independent samples t-test found these category bounds did not differ statistically [$t(27.916) = 0.287$, $p = 0.776$].

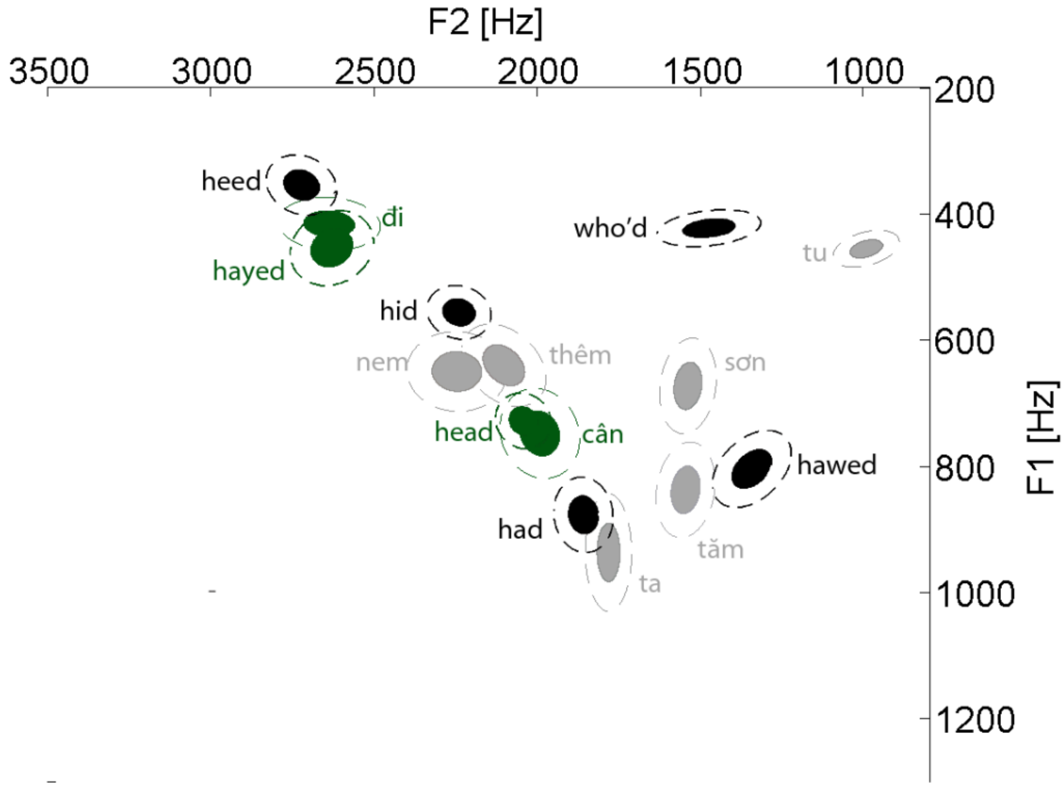


Figure 21: Phonetic category assimilation in the common phonological space for female Vietnamese talkers (n=17).

Black represents English vowels. Grey represents Vietnamese vowels. Green represents vowels likely to have experienced phonetic category assimilation. The center of each ellipse represents the mean F1 and F2 frequencies. The solid and dashed ellipses represent one and two standard deviations, respectively.

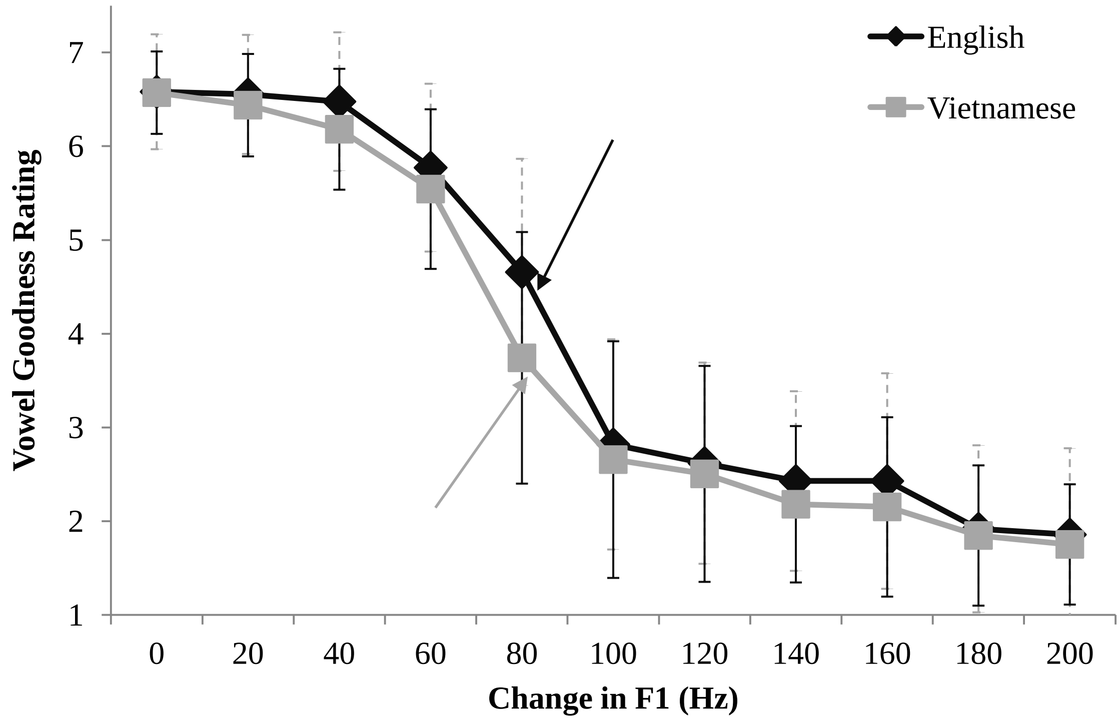


Figure 22: Mean vowel goodness ratings.

Black diamonds represent English talkers. Grey squares represent Vietnamese talkers. Arrows indicate location of vowel category bounds.

3.1 Compensation comparisons

Individuals vary in their absolute F1 for a given vowel. Therefore F1 productions were normalized to allow comparison across individuals. For each individual, the change in F1 production was normalized by subtracting the average F1 of the Baseline phase from the average F1 of each step (6 utterances per step) in the Ramp phase or the 19 utterances from the Hold phase. In the following ANOVA analyses, absolute values of the normalized data from each participant were used because of the different signs of the positive and negative feedback shifts. For the repeated measures ANOVA analyses, within-subject factors were shift size (20, 40, 60, ...200 Hz) and direction (positive and negative), and the between-subject factor was group (Vietnamese and English talkers). Max and Onghena (1999) reported that the sphericity assumption in repeated measures ANOVA are usually violated and even if the data passes the spehericity test, sphericity assumed values should not be reported. As a result, Greenhouse-Geisser (Greenhouse & Geisser, 1959) values were reported in the current study. ANOVA analyses were performed to determine the effects of direction and shift size. In the ANOVA analyses, differences between groups were not likely to be shown because utterances during the Baseline, Ramp, and Hold phases were averaged across trials into one mean value per phase or Ramp step. As a result, an increase of variance would occur in the ANOVA data, which may have diluted group differences and the ANOVA analyses may have difficulties evaluating significant group differences. Instead, a Sign test was used to evaluate group differences for compensation of English vowels /ɪ/ and /æ/. For each group, a set of average utterances were calculated across participants, where each utterance was an average of all the participants. The number of average utterances where one group's compensation exceeded the other was calculated from speech compensation threshold to the end of the Ramp phase for the Sign test. This was repeatedly separately for the Hold phase. Group speech compensation threshold was defined as the point in the Ramp phase when the average change in F1 production was two standard deviations from average Baseline for each group (refer to ++ symbols in Figures 23, 24 and 25).

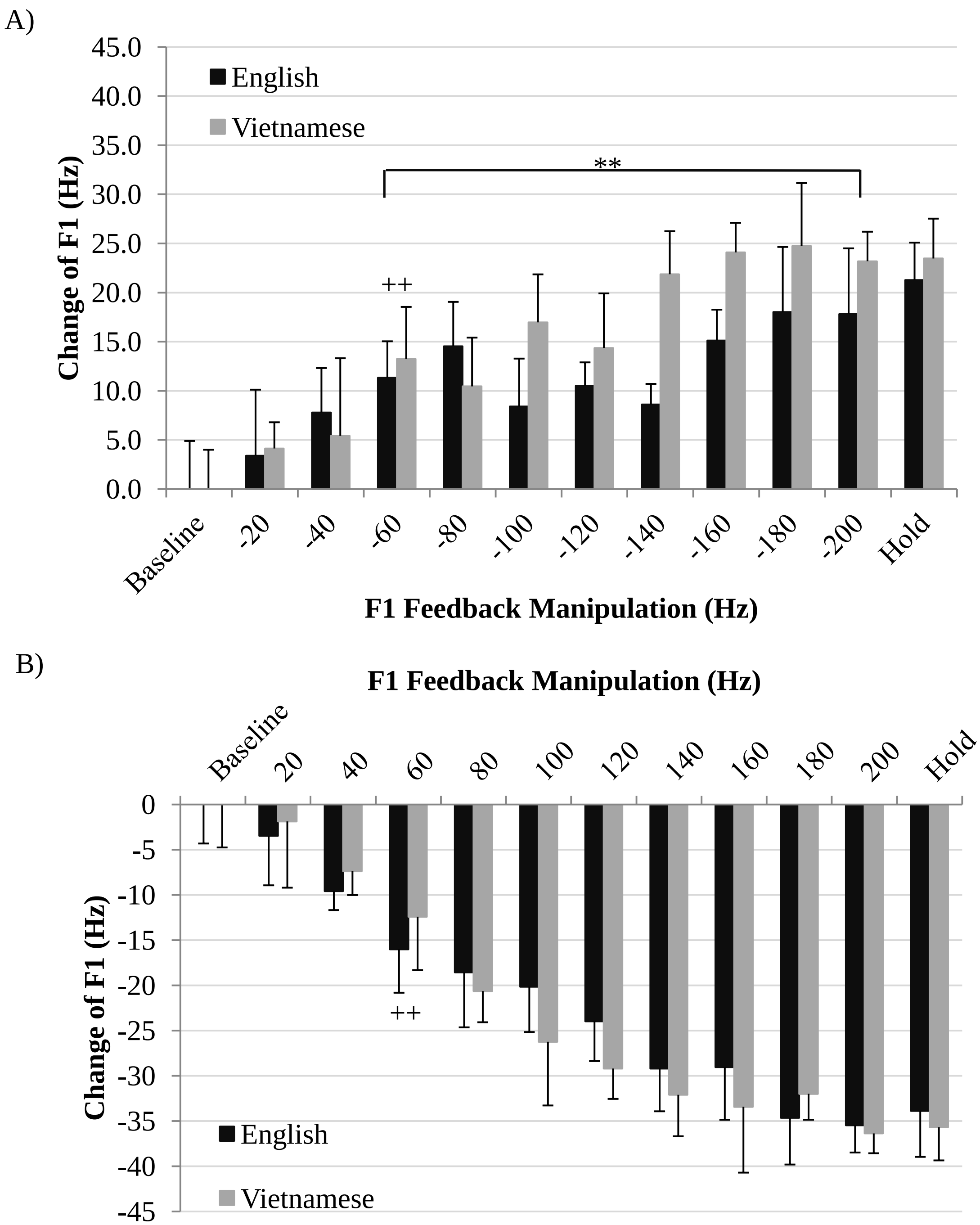
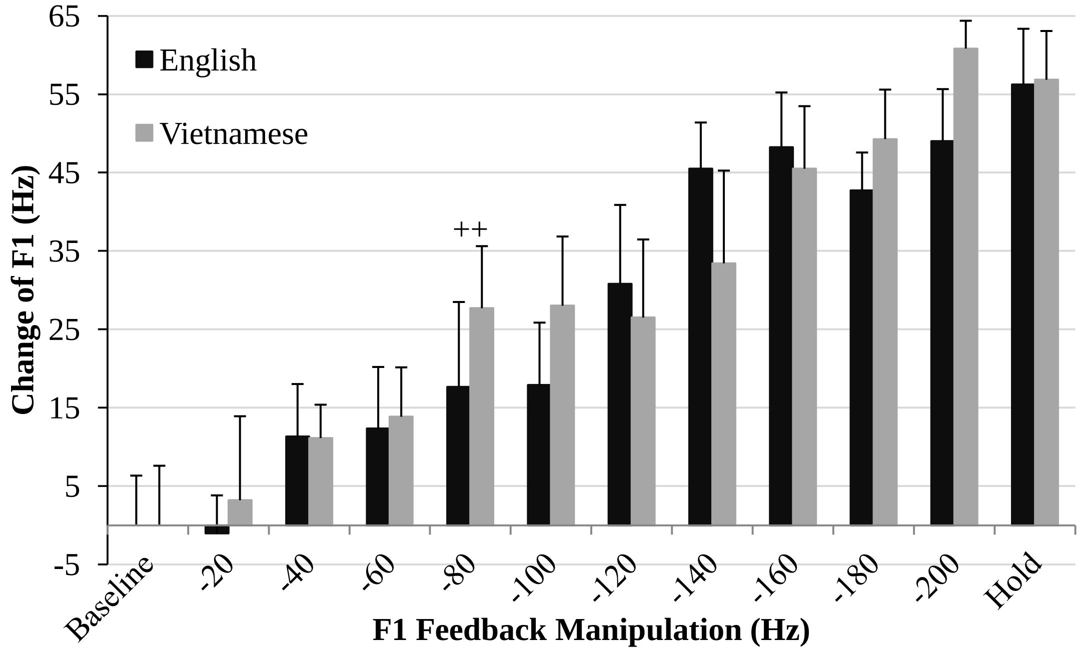


Figure 23: Average normalized F1 results for English /ɪ/ in ‘hid’ across trials for English and Vietnamese speakers.

A) Negative condition B) Positive condition. Ramp steps (±20, ±40 … ±200) indicate averaged utterances for each shift size. Black represents English talkers. Grey represents Vietnamese talkers. Error bars represent one standard deviation. ++ represents speech compensation threshold. ** represents significant difference between groups.

A)



B)

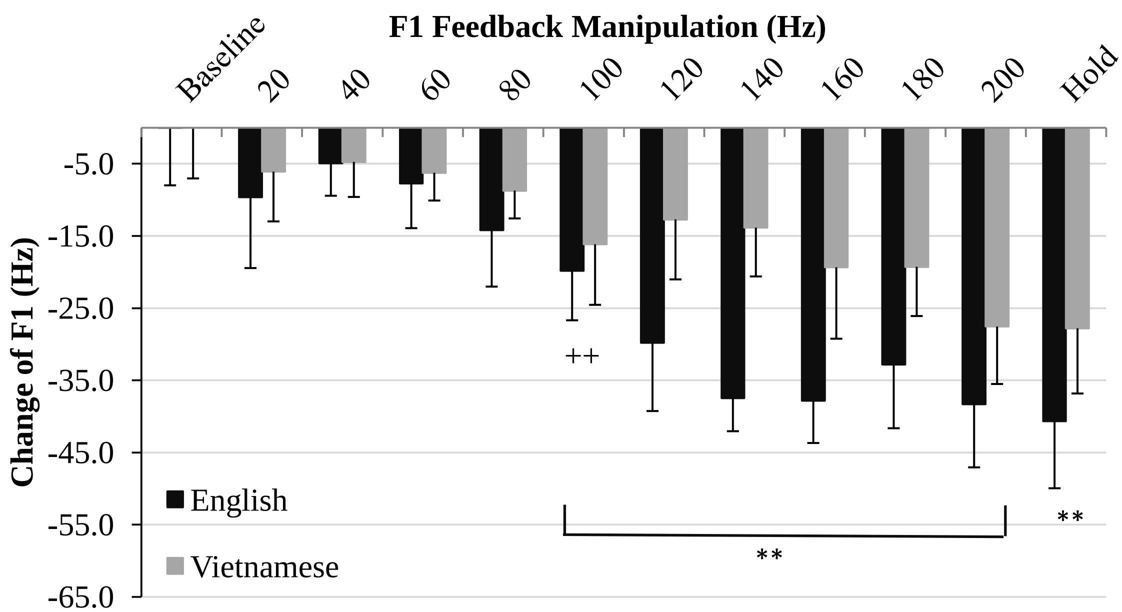


Figure 24: Average normalized F1 results for English /æ/ in ‘had’ across trials for English and Vietnamese talkers.

A) Negative condition B) Positive condition. Ramp steps (±20, ±40 … ±200) indicate averaged utterances for each shift size Black represents English talkers. Grey represents Vietnamese talkers. Error bars represent one standard deviation. ++ represents speech compensation threshold. ** represents significant difference between groups.

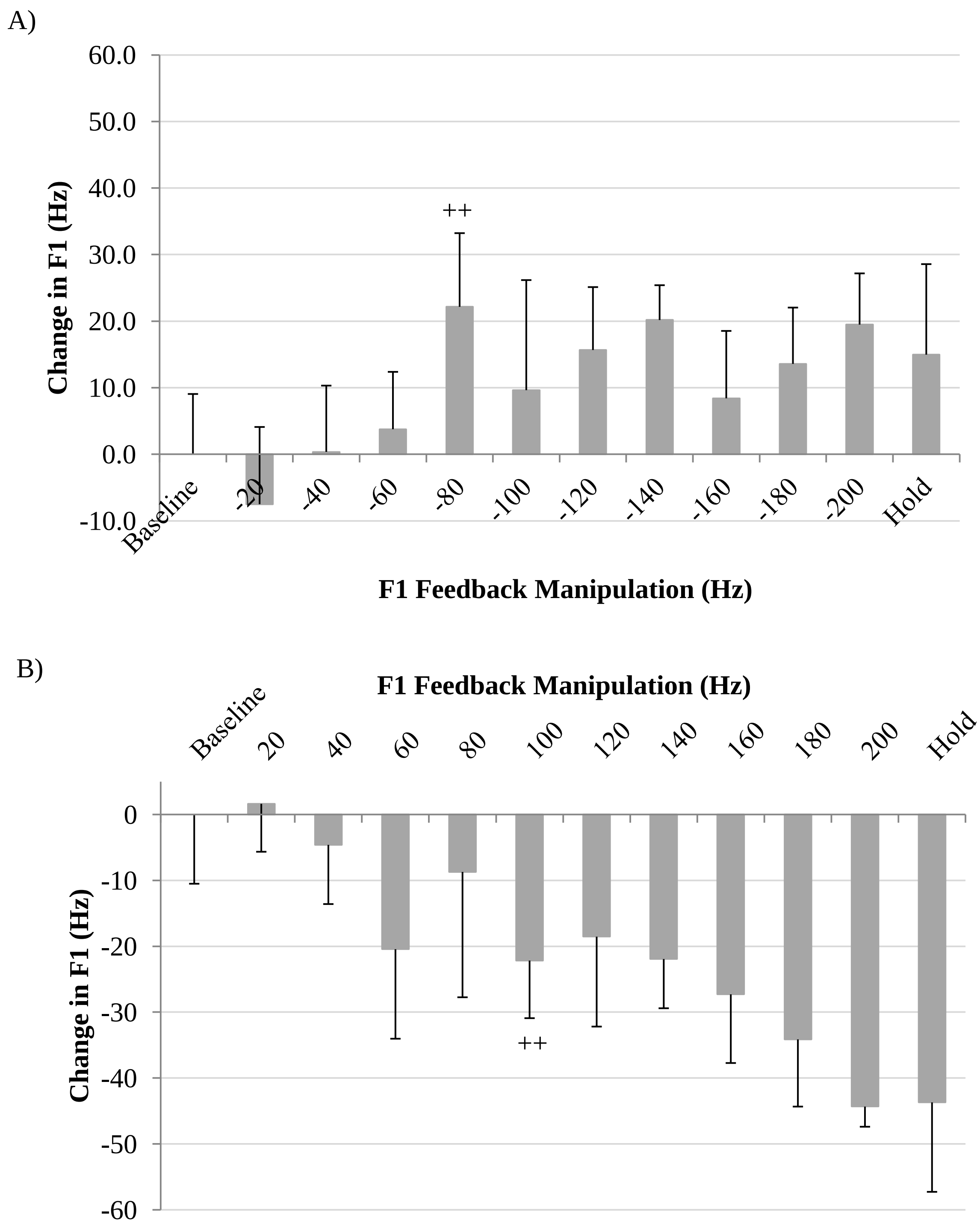


Figure 25: Average normalized F1 results for Vietnamese /ɐ/ in 'tăm' across trials for Vietnamese speakers.

A) Negative condition B) Positive condition. Ramp steps (±20, ±40 … ±200) indicate averaged utterances for each shift size Error bars represent one standard deviation. ++ represents speech compensation threshold.

3.1.1 Speech compensation for English /ɪ/ in ‘hid’

The results for the F1-negative condition for the English vowel /ɪ/ are plotted in Figure 23A. The results for the F1-positive condition for English /ɪ/ are plotted in Figure 23B. Differences in the magnitudes of compensations between the negative and positive shifts were observed. Talkers compensated more in the positive shift than in the negative shift (ANOVA; $F[1,40] = 6.604$, $p < 0.014$, $\eta^2 = 0.142$). Group differences were observed only in the negative shift during the Ramp phase, where Vietnamese talkers compensated more than English talkers (Sign test; $Z = -3.753$, $p < 0.0002$). During the Hold phase, where the F1 shift was held constant at ± 200 Hz, the two groups did not differ in their compensation (Sign test; positive shift: $Z = 1.842$, $p = 0.359$; negative shift: $Z = -0.688$, $p = 0.648$). Group speech compensation thresholds were found to be the same for English and Vietnamese talkers. The speech compensation threshold was at ±60 Hz F1 manipulation during the Ramp phase for both shift conditions.

3.1.2 Speech compensation for English /æ/ in ‘had’

The results for the F1-negative condition for the English vowel /æ/ are plotted in Figure 24A. The results for the F1-positive condition for English /æ/ are plotted in Figure 24B. Differences in the magnitudes of compensations between the negative and positive shifts were observed. Talkers compensated more in the negative shift than in the positive shift (ANOVA; $F[1,40] = 6.116$, $p < 0.018$, $\eta^2 = 0.133$). Group differences were observed only in the positive shift during the Ramp phase, English talkers compensated more than Vietnamese talkers (Sign test; $Z = -4.000$, $p < 0.0002$). During the Hold phase, in which the F1 shift was held constant at ± 200 Hz, group differences were observed only in the positive shift, where English talkers compensated more than Vietnamese talkers (Sign test; $Z = -2.982$, $p < 0.004$). Group speech compensation thresholds were found to be the same for English and Vietnamese talkers. The speech compensation threshold was at ±60 Hz F1 manipulation during the Ramp phase for both shift conditions.

3.1.3 Speech compensation for Vietnamese /ɐ/ in 'tăm'

The Vietnamese talkers showed compensation when presented with altered auditory feedback for the Vietnamese vowel /ɐ/. The results for the F1-negative and positive conditions for Vietnamese /ɐ/ are plotted in Figure 25A and 25B respectively. An asymmetry of compensation occurred, in which the positive shift had greater compensation than the negative shift (ANOVA; $F[1,20] = 5.683$, $p < 0.027$, $\eta^2 = 0.221$). Speech compensation threshold for the negative shift was at -80 Hz during F1 manipulation and for the positive condition it was at 100 Hz.

3.1.4 General patterns in speech compensations for English /ɪ, æ/ in 'hid, had' and Vietnamese /ɐ/ in 'tăm'

All experimental conditions showed a significant effect for shift size, in which an increase of shift size resulted in greater compensation (ANOVA; /ɪ/: $F[3.256, 130.224] = 27.809$, $p < 0.001$, $\eta^2 = 0.410$; /æ/: $F[4.727,189.094] = 22.667$, $p < 0.001$, $\eta^2 = 0.362$; /ɐ/: $F[5.096,101.924] = 14.391$, $p < 0.001$, $\eta^2 = 0.418$).

Previous studies, such as MacDonald et al. (2010, 2011) have performed correlations with vowel distances and magnitude of speech compensations. These correlations were also replicated in the present study. Correlations between vowel distances (see sections 3.1.1 and 3.1.2) and magnitude of speech compensation did not find significant effects for any vowel conditions, except for Vietnamese talkers with /ɪ/-positive [Pearson; $r(15) = -0.513$, $p < 0.035$].

3.2 Interactions between perception and production

It was predicted that the F1 discrimination threshold and vowel goodness may be related to compensation results in altered auditory feedback. Comparisons across perceptual measures and speech compensation thresholds were completed to determine if these measures were similar or different. Data analyses reported above found that English and Vietnamese talkers did not differ in their F1 discrimination threshold, vowel category bounds, and speech compensation thresholds; therefore the English and Vietnamese talkers were treated as one group. Multiple paired t-tests were performed to compare F1 discrimination, vowel category bounds, and speech compensation thresholds. A paired t-test found that the F1 discrimination threshold and vowel category bound were significantly different [$t(21) = -15.771$, $p < 0.001$], in which the F1 discrimination threshold was smaller. This suggested that F1 discrimination threshold and vowel category bounds may be related to speech compensation differently. Therefore, paired t-tests between F1 discrimination threshold or vowel category bounds and speech compensation threshold were performed. F1 discrimination and speech compensation thresholds for each participant were significantly different [$t(21) = -6.190$, $p < 0.001$]. In contrast, the vowel category bound and speech compensation threshold for each participant did not differ statistically [$t(21) = -1.701$, $p = 0.104$]. This result suggested that the vowel goodness ratings for each different /ɪ/-like sound and the magnitude of speech compensation at each Ramp step may have a relationship.

Corresponding speech compensation values for each step during the Ramp phase of /ɪ/ during the positive shift were correlated with vowel goodness ratings for each group (Figure 26). Vietnamese and English talkers were treated as separate groups because they significantly differed in their vowel goodness ratings (see section 3.2.2). Pearson correlations found the linear relationship between speech compensation and vowel goodness ratings to be robust and significant [English: $r(9) = 0.940$, $p < 0.001$; Vietnamese: $r(9) = 0.989$, $p < 0.001$]. Overall, high goodness ratings corresponded to low compensations, and lower goodness ratings corresponded to greater speech compensations.

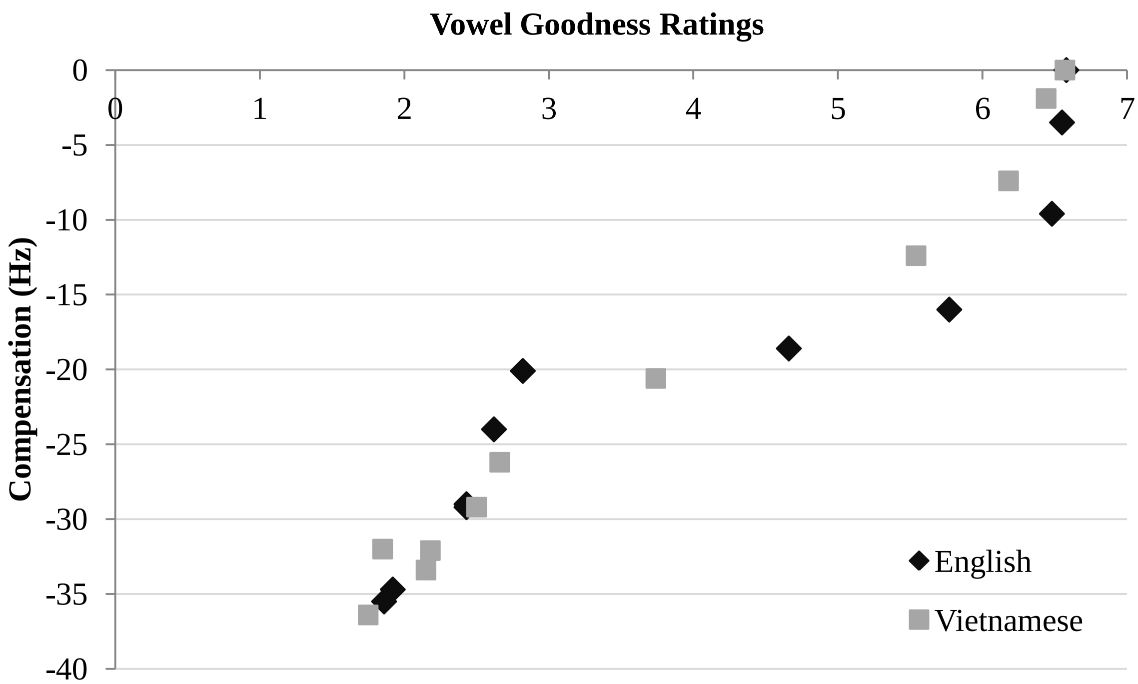


Figure 26: Relationship between vowel goodness ratings and speech compensation for /ɪ/ in 'hid' on a continuum towards /ɛ/ in 'head'.

Black diamonds represent English talkers. Grey squares represent Vietnamese talkers.

Chapter 4

4 Discussion

The purpose of the current study was to examine mechanisms underlying the use of auditory feedback in the control of speech production. The approach used here was to elicit compensatory speech production responses to altered auditory feedback in native talkers of English and Vietnamese. If native language plays a role in how auditory feedback is used then differences in compensation would be expected between the two groups. Both groups performed psychoacoustic listening tasks to determine whether they had perceptual differences. Talkers produced vowels from across the English vowel space to evaluate production differences. Both groups of talkers were given perturbations in feedback where F1 of English vowels /ɪ/ and /æ/ were increased or decreased to investigate whether they made different use of auditory feedback. The following sections will discuss the results and relate them to what is previously known from the literature.

4.1 Vowel productions

The Vietnamese talkers reported that they are fluent in English, and that English is presently their dominant language. They are fluent and dominant in English because their environments, such as school, may demand high English use. Vietnamese talkers also reported that they mainly use Vietnamese when they are speaking to family. This is congruent with language dominance research by Grosjean (1982, 1989, 1997) and studies by Flege and colleagues (Flege, 2002; Flege et al., 1995ab; Yeni-Komshian et al., 2000) where they reported that early bilinguals are dominant in their L2 because of high-use of L2 in their daily surroundings.

Even though the Vietnamese talkers are English dominant, the influence of L1 Vietnamese on L2 English was evident. Productions of the English vowels across the vowel space were similar between the two groups, except for /u/, in which Vietnamese talkers had a higher F1 than English talkers. Vietnamese talkers gave slightly lower ratings to the different exemplars of /ɪ/-positive during the vowel goodness test in comparison to English talkers. As well, group differences occurred in some of the conditions during altered auditory feedback, such as /æ/-positive and /ɪ/-negative. These results are consistent with other language acquisition research where it is suggested that L1 and L2 may interfere and interact with each other, leading to differences in bilingual and monolingual speakers (Grosjean, 1982, 1989, 19997; Flege, 1995, 2002; Yeni-Komshian et al., 2000). This suggests that there may always be differences between fluent L2 English speakers and L1 English speakers.

Interactions of L1 Vietnamese and L2 English phonemes may have occurred within the CPS of the Vietnamese bilingual talkers. The analysis of F1 and F2 between English and Vietnamese vowels found that Vietnamese /i/ and English /e/, as well, Vietnamese /ʌ/ and English /ɛ/ were similar (see Figure 21). This suggest that PCA (Flege, 1995) may have occurred between these two sets of vowels. It may be possible that the Vietnamese talkers were using their L1 category of /i/ and /ʌ/ to produce and perceive English /e/ and /ɛ/ sounds, respectively. In contrast, the analysis found that English and Vietnamese differed in their production of English /u/. This suggested that the English /u/ category in Vietnamese talkers may have interacted with a Vietnamese vowel category, Vietnamese /u/. To maintain two distinct categories, English /u/ and Vietnamese /u/ in the bilingual talkers may have deflected away from each other, leading to differences in the production of English /u/ in comparison to English monolinguals. This suggests that PCD (Flege, 1995) may have occurred. These results support Grosjean's theory of cross-language interference (see section 1.3.3.1; 1982, 1989, 1997), in which he states that bilinguals and monolinguals would always be different from each other. These findings are suggestive and other future experiments are needed to determine the presence of PCA, PCD, and cross-language interference in Vietnamese bilingual talkers.

4.2 Factors underlying auditory feedback

Speech compensation occurred in Vietnamese talkers for Vietnamese vowel /ɐ/ during altered auditory feedback. Speech compensation opposed the F1 manipulation, greater compensation occurred with increasing F1 shift size, and compensation was smaller than the F1 manipulation. These trends were similar to other pitch and formant-manipulated studies with English talkers (Jones & Munhall, 2000; MacDonald et al.,2010; Munhall et al., 2009; Purcell & Munhall, 2006a) and other cross-language studies, such as Mandarin (Jones & Munhall, 2002) and Japanese (Mitsuya et al., 2011). Other speech compensation results from English and Vietnamese talkers for English vowels /ɪ, æ/ from the current study also had an effect of shift size and compensations in the opposite direction of the F1 manipulation. This suggests that compensation to altered auditory feedback may be universal across languages, where talkers tend to compensate for formant errors in the opposite direction.

Two factors were evaluated that may play a role in the use of auditory feedback in the correction of speech production. The first is the NRV perspective (Polka & Bohn, 2003) in which shifts towards peripheral vowels may be easier to discriminate and thus tend to elicit larger responses to feedback manipulation. The second is differences in the weight of auditory cues and somatosensory cues, which may be based on vowel location. If feedback manipulation is outside the vowel space, robust somatosensory cues may be weighed more heavily, eliciting smaller compensations. The following will discuss where these mechanisms may have been invoked in the present study.

4.2.1 Natural Referent Vowel (NRV) perspective

Vietnamese talkers had an asymmetry in compensation for Vietnamese /ɐ/, in which the magnitude of compensation was greater in the positive shift than in the negative shift. The location of /ɐ/ in the Vietnamese talkers' CPS may have influenced this asymmetry through the NRV effect. The negative shift of /ɐ/ moved the feedback towards the central area of the CPS and away from the peripheral vowel /a/ (see Figure 20). However, in the positive shift, the feedback was towards the peripheral vowel /a/. The NRV perspective would suggest that changes in the feedback would be more salient in the positive shift than in the negative shift, which corresponded with greater compensation. The study by Mitsuya et al. (2011) also suggested an influence of the NRV in Japanese talkers. They reported that in the positive-shift condition, where the feedback was being shifted away from a peripheral vowel in Japanese talkers, there were smaller compensations. As well, an influence of NRV may have occurred in other formant-manipulated studies that have used /ɛ/, a central vowel (MacDonald et al., 2010; Mitsuya et al., 2011; Munhall et al., 2009; Purcell & Munhall, 2006a). These studies found symmetrical compensations in the negative and positive shifts. Both shift directions were towards functionally more peripheral reference vowels /ɪ/ and /æ/ (Mitsuya, et al., 2011). Polka and Bohn (2003) suggested that changes toward the periphery of CPS were more salient because vowels in the periphery area may be more frequent, familiar, intense and information rich. As well the periphery vowels are located at the limits of the articulatory/acoustic space of F1 and F2. Because they are at the limits, studies have found that there were less acoustic overlap of point vowels with another vowel category (Polka and Bohn, 2003) and there were lower confusion rates than other central vowels (Peterson & Barney, 1952). Further investigation is needed to determine the influence of NRV on compensation, in which direction of shift based on vowel location may influence the saliency of feedback errors.

4.2.2 Vowel space boundaries

There were differences in speech compensation between groups for manipulation of vowel /æ/ in the positive shift direction. English talkers compensated more than Vietnamese talkers for both the Ramp and Hold phases of this shift direction. The location of /æ/ in the English vowel space and in the Vietnamese CPS may have affected these compensation results. In the English vowel space (see Figure 16A), /æ/ is a point vowel; a shift in the negative direction would be within the vowel space, but a shift in the positive direction would be outside the vowel space. In contrast, a shift in the negative and positive direction would have been within the Vietnamese CPS (see Figure 20). The NRV perspective predicted that Vietnamese talkers would compensate more than English talkers in /æ/-positive, because English /æ/ was perturbed towards Vietnamese /a/, a peripheral vowel; however, this did not occur. This suggest that the NRV is only applicable for specific phonemes within a language or that the English talkers were more strongly influenced by the shift outside their vowel space.

Asymmetries in compensation between the negative and positive shifts of English /ɪ/ and /æ/ occurred in the present study for both groups. In English /ɪ/, there was less compensation in the negative condition; whereas, English /æ/ had less compensation in the positive condition. In these conditions, the feedback shifts were towards the extremes of the English vowel space. Previous studies have found that when the feedback perturbation was towards the extremes of the vowel space, less compensation occurred (Mitsuya, et al., 2011; Purcell et al., 2011). This asymmetry in compensation when perturbed within and outside the vowel space may be influenced by how the system weighs somatosensory and auditory feedback. When the perturbation is outside the vowel space, the system may rely more strongly on somatosensory cues than auditory cues (MacDonald et al., 2010; Purcell et al., 2011). The system may expect that when an auditory signal is valid, that robust somatosensory feedback would be present. However, these auditory feedback perturbations of 200 Hz towards the extremes of the English vowel space may be treated as unnatural since the somatosensory cues were unmanipulated and therefore contradictory. The system may have reduced the weight of the auditory cues in the control of speech production relative to the more central shifts. These results suggest that the system may weigh feedback cues based on the vowel location.

Results from the current study suggest an importance of vowel location within the talkers' vowel space for speech compensation in perturbed auditory feedback. Two mechanisms involving vowel location were considered given the current results: the NRV perspective and the balance of somatosensory and auditory cues. Future studies would need to be designed to tease apart the relative contributions (if any) from these two mechanisms. An additional interesting result came forth, during which in the negative shift for English /ɪ/ in the Ramp phase, Vietnamese talkers compensated more than English talkers; however, in the Hold phase, the two groups did not differ. This suggests that the Vietnamese talkers' system is reacting differently to the dynamic manipulation than the English talkers' system. Further investigation is needed to explore these complex results.

4.3 Perceptual measures

The auditory feedback system may be responding to changes in F1 that are substantially greater than the F1 discrimination threshold. Talkers' normal productions of vowels vary (Munhall et al., 2009) and as a result the system may not micro-manage errors in production, if these errors are within an acceptable range. Purcell and Munhall (2006b) found that the F1 speech compensation threshold was approximately two-to-three standard deviations beyond the mean of normal F1 productions. It appears that the system expects variability in production and will only correct for errors when production is substantially different. This may explain why the talkers' did not start to compensate at the F1 discrimination threshold but at a later point during feedback manipulation. Another possible reason to explain the differences between the thresholds is that the thresholds are two different measures. The 2AFC test is a psychoacoustic test during which the participant is listening to differences. In contrast, the feedback control system is both listening and reacting to differences in speech production as the person is talking. The F1 discrimination threshold does not directly inform us about the threshold at which the system begins to correct for speech errors, but rather provides a lower bound for errors in the frequencies of the vowel category.

Production variability of a vowel, such as “hid”, differs between speakers. Individuals are able to perceive this variability as one vowel category. However, certain exemplars are perceived to be better prototypes. The vowel goodness test was used to categorize the variability of “hid” on a continuum towards /ɛ/ between a prototype and non-prototype. Results from the vowel goodness ratings test indicated that the Vietnamese bilingual talkers gave lower goodness ratings to the various exemplars of “hid” than the English monolingual talkers. The Vietnamese bilingual talkers may have given lower ratings because L2 learners may emphasize and prefer the prototypical version of a vowel category. As a result, other versions of the vowel that is not prototypical would be considered a little less good. Despite these group differences, a common pattern occurred in which English and Vietnamese talkers gave higher ratings to /ɪ/-like sounds with small F1 changes and lower ratings to /ɪ/-like sounds with large F1 changes. The relationship between different /ɪ/-like sounds and vowel goodness ratings was non-linear (see Figure

22). This suggests that the perception of a bad and good sound of “hid” was categorical, where the different exemplars were perceived as either “good” or “bad”. The threshold between a prototypical and non-prototypical version of “hid”, also known as the vowel category bound, was calculated to be 80 Hz. These findings were similar to studies by Kuhl and colleagues. Kuhl (1991) found that prototypical /i/-like sounds had higher goodness ratings and non-prototypical /i/-like sounds had low goodness ratings across participants. This shows that the exemplars in a vowel category are not all perceptually equal. The similarities between talkers of different language backgrounds suggest that representations of vowel categories are consistent across individuals, where individuals perceive exemplars within a vowel category similarly despite language background differences.

4.4 Links between production and perception

The auditory feedback system is able to correct for small errors above the F1 discrimination threshold. However, significant compensations occurred at the compensation threshold, the limit of normal production variability. Results found the compensation threshold numerically approached the perceptual vowel category bound and did not differ from it statistically. The perceptual vowel category bound is the limit of normal variability for a good exemplar. Within a vowel category, different exemplars of /ɪ/ were given different vowel goodness ratings and these corresponded to differences in the magnitude of compensation. Before the vowel category bound, /ɪ/-like exemplars were given high goodness ratings and had numerically small compensations that occurred in the opposite direction of the F1 manipulation (see Figure 26). This showed that the system detected a change in F1 feedback that was different from the prototype and responded with numerically small compensations. However, a significant change in compensation occurred when F1 manipulation was close to the vowel category bound. Significant speech compensations occurred when the feedback error was similar to those of non-prototypical /ɪ/-like sounds. If the feedback does not sound like the intended vowel, the system may be less tolerant of incongruence between feedback and production, leading to greater compensations. Even though the auditory system was able to detect changes in F1 of 30 Hz (the F1 discrimination threshold), these sounds may have been categorized as good variants of “hid”. This may explain why significant speech compensation did not start at the F1 discrimination threshold because the error at 30 Hz shift may have been processed as good feedback for the /ɪ/-category. Therefore, there seems to be a relationship between the limits of normal production variability and the perceptual vowel category bounds. Further investigation is needed to explore the influence of vowel category bounds on speech compensation in altered auditory feedback.

Furthermore, similarities and differences between Vietnamese and English talkers in speech compensation may have been dependent on how their systems processed the F1 manipulation. Vietnamese and English talkers had similar compensations at each step during the Ramp phase for /ɪ/-positive and both groups had a speech compensation threshold of 60 Hz. Perceptual results found that Vietnamese and English talkers categorized the different exemplars of /ɪ/-positive similarly (albeit shifted slightly). As well, the F1 discrimination thresholds and vowel category bounds were not significantly different between the two groups. This suggests that compensation between the groups during /ɪ/-positive was similar because their systems processed the F1 manipulations in a similar manner. Speech compensation during the Ramp phase in /æ/-negative also did not find group differences. Results from /ɪ/-positive suggest that the two groups also processed the /æ/-negative F1 manipulations in a similar manner. In contrast, /ɪ/-negative and /æ/-positive had group differences in compensation during the Ramp phase. Perhaps the two groups processed the F1 manipulations in a different manner, possibly due to local differences in their phonological spaces. For instance, as discussed above, /æ/-positive was shifted outside the English vowel space, whereas the manipulation remained inside the CPS of the Vietnamese talkers, potentially leading to group differences in compensation. Since the perceptual tasks were limited only to /ɪ/-positive, further perceptual tasks are needed on other vowel conditions to determine the relationship between speech compensation and how F1 perturbations are processed by the system.

The feedback system may be more sensitive to the intended vowel category and less sensitive to other neighbouring categories. Results from the present study found negative and positive distances to neighbouring vowel categories of the manipulated vowels were different. For instance, the nearest vowel category in the negative direction for /ɪ/ was greater than in the positive direction. Correspondingly, the positive shift condition for /ɪ/ had greater compensation than the negative condition. This may have suggested that density of the vowel space influences compensation results, where smaller distances would result in greater compensation. However, the correlations between distances of adjacent vowel categories and magnitude of compensation were not significant. This is consistent with studies by MacDonald et al. (2010, 2011) who suggested that neighbouring vowel categories do not influence compensation results. Instead, the feedback system may be attuned to the intended vowel category only, where significant compensation onset occurs near the respective vowel category bound. For instance, during altered auditory feedback, /ɪ/ was shifted by 200 Hz in the positive direction into the /ɛ/ category, where the feedback would have sounded like “head”. The system may not have processed the feedback as belonging to the /ɛ/ category, but rather as an error within the /ɪ/ category. The phonological mediation of auditory feedback may be sensitive to variability within a specific category and not to the interactions between categories. Further investigations are needed to determine the sensitivity of the auditory feedback to within category variability.

4.5 Concluding remarks, limitations and future work

In summary, the purpose of the study was to examine the underlying mechanisms of the auditory feedback system using Vietnamese and English talkers in response to feedback perturbations. Participants consisted of native Vietnamese speakers with ESL and English monolingual speakers. F1 discrimination thresholds, vowel goodness ratings, vowel category bounds, vowel spaces, and speech compensations during altered auditory feedback were collected from all participants. Similarities between the Vietnamese bilinguals and English monolinguals in perceptual measures, vowel space plots, and some speech compensation results confirm the Vietnamese talkers self-report of extensive English experience. Speech compensation during perturbed auditory feedback occurred in English and Vietnamese vowels suggesting that the underlying mechanisms are universal.

Despite these similarities, there were differences in speech compensation for some vowel conditions, which may have occurred because of difference in the target vowel location in each group's vowel space. Location in the vowel space may dictate the weight of different feedback cues that the system may use to detect errors. The results suggest that when F1 was perturbed within the vowel space, the system may rely on auditory cues and compensate more. In contrast, when it was perturbed outside the vowel space, somatosensory cues may be more heavily used leading to less compensation. For middle vowels, if the feedback is shifted towards a peripheral vowel, compensation might be greater because changes in feedback are more salient, as described by the NRV perspective. Speech compensation may also be influenced by the boundaries of normal production and category bound of a target vowel. Significant compensation may occur when the system detects production variability that is atypical and when the feedback is non-prototypical to its intended category. Overall, results from the current study suggest that the feedback system is sensitive to formant frequency changes and these may have underlying phonological mediations.

There are several limitations that can be attributed to the current study that may have influenced the results. One limitation could be participant sampling. Vietnamese talkers used in the study were a mixture of the three dialects: Northern, Central, and Southern. Even though speech compensation results were normalized, dialect issues may have arisen. Another limitation of the study could be that the perceptual tests were only conducted for /ɪ/-positive. Perceptual measures were only conducted on one vowel condition because of time limitations and concern for participant fatigue. As a result, implications on speech production and perception may be specific to /ɪ/-positive.

Future research should focus on exploring altered auditory feedback with other vowels and languages. The current study was limited to two English vowels and one Vietnamese vowel. Future research should compare other English and Vietnamese vowels. For instance, the vowel space analysis found that English /ɛ/ and Vietnamese /ʌ/ had similar F1 and F2 values and English /u/ was produced differently by English and Vietnamese speakers. It would be interesting to perform altered auditory feedback with these three vowels. Comparisons with others languages, such as Mandarin could also be done. It would be interesting to extend the work on pitch-shifted auditory feedback in Mandarin (Jones & Munhall, 2002; Xu, Larson, Bauer, & Hain, 2004) to formant-shifted auditory feedback experiments. In addition, future studies should include perceptual measures for all vowel conditions and using stimuli that are more ecologically valid. For example, during the 2AFC test, stimuli may use samples from the talkers' own voice.

Research using altered auditory feedback may offer insight on how the brain regulates speech production and how it detects and corrects for speech errors. Reactions to speech errors by the system occur unconsciously and automatically. Further research is needed to understand how the system reacts to errors in auditory feedback and how the system uses feedback cues to detect errors. This understanding may benefit individuals who are learning languages, hearing-impaired, and/or those with speech disorders. Language teachers, therapists, and others could teach these individuals how to attend most effectively to their own voice and be conscious of the errors in their pronunciations. Research using altered auditory feedback may lead to benefits in language fluency, proficiency, and rehabilitation.

References

Bauer, J., Mittal, J., Larson, C., & Hain, T. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voiceamplitude. *Journal of the Acoustical Society of America, 119*(4), 2363-1371. doi: 10.1121/1.2173513

Bever, T. (1981). Normal acquisition processes explain the critical period for language learning. In K. Diller (Ed.), *Individual Differences and Universal in Language Learning Aptitude* (pp. 176-198). Rowley, MA: Newbury House.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language, 68*, 706-755. doi: 10.2307/416851

Bishop, D. (n.d.). Dinosaur [computer software]. Oxford University.

Boberg, C. (2005). The Canadian shift in Montreal. *Language Variation and Change, 17*, 133-154. doi: 10.1017/S0954395050064

Bohn, O.-S., & Flege, J. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition, 14*, 131-158. doi: 10.1017/D0272263100010792

Bohn, O.-S., & Polka, L. (2001). Target spectral, dynamic spectral, and duration cues in infant perception of German vowels. *Journal of the Acoustical Society of America, 110*, 504-515. doi: 10.1121/1.1380415

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition, 19*, 447-465. doi: 10.1017/S0272263197004026

Borden, G., Harris, K., & Raphael, L. (1994). *Speech Science Primer: Physiology, Acoustics, and Perception of Speech.* (3rd ed.). Baltimore, MD: Williams & Wilkins.

Bradlow, A. (2010). *Archive of L1 & L2 scripted and spontaneous transcriptions and recordings (ALLSSTAR).* Retrieved from http://groups.linguistics.northwestern.edu/speech_comm_group/allsstar/index.html

Burnett, T., Freedland, M., Larson, C., & Hain, T. (1998). Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America, 103*(6), 3153-3161. doi: 10.1121/1.423073

Chen, C., Li, H., Shen, L., & Fu, G. (2001). Recognize tone languages using pitch information on the main vowel of each syllable. *Acoustics, Speech, and Signal Processing, 1*, 61-64. doi: 10.1109/ICASSP.2001.940767

Clarke, S., Elms, F., & Youssef, A. (1995). The third dialect of English: Some Canadian evidence. *Language Variation and Change, 7*, 209-228. doi: 10.1017/S0954394500000995

Cooke, D.M. (n.d.). *Occupational Therapy Adult Perceptual Screening Test (OT-APST).* Retrieved from http://www.functionforlife.com.au/

DeKeyser, R. (1990). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*, 499-534. doi: 10.1007/s11155-006-9021-6

Edmondson, J. (2009). Vietnamese. In K. Brown, & S. Ogilvie (Eds.). *Concise Encyclopedia of Languages of the World* (pp. 432-436). Oxford, UK: Elsevier Ltd.

Flege, J. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced a second language. *Journal of the Acoustical Society of America, 89*, 395-411. doi: 10.1121/1.400473

Flege, J. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An Integrated View of Language Development: Papers in Honor of Hennig Wode.* Trier, Germany: Wissenchaftlicher Verlag Trier.

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

Flege, J. (1992). Speech learning in a second language. In C. Ferguson, L. Menn, & C. Stoel-Grammon (Eds.), *Phonological Development: Models, Research, and Implications* (pp. 565-604). Timonium, MD: York Press.

Flege, J. (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. Retrieve from http://jimflege.com/files/Flege_new_similar_JP_1987.pdf

Flege, J., & Eefting, W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics, 15*, 67-83. Retrieve from http://jimflege.com/files/Flege_Eefting_English_stops_JP_1987.pdf

Flege, J., MacKay, I., & Meador, D. (1999a). Native Italian speakers' production and perception of English vowels. *Journal of the Acoustical Society of America, 106,* 2973-2987. doi: 10.1121/1.428116

Flege, J., Munro, M., & MacKay, I. (1995a). Factors affecting strength of perceived foreign accent in a seocnd language. *Journal of the Acoustical Society of America, 97*, 3125-3134. doi: 10.1121/1.413041

Flege, J., Munro, M., & MacKay, I. (1995b). The effect of age of second language learning on the production of English consonants. *Speech Communication, 16,* 1-26. doi: 10.1016/0167-6393(94)00044-B

Flege, J., Schirru, C., & MacKay, I. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication, 40*, 467-491. doi: 10.1016/S0167-6393(02)00128-0

Flege, J., Yeni-Komshian, G., & Liu, H. (1999b). Age constraints on second language acquisition. *Journal of Memory & Language, 41*, 78-104. doi: 10.1006/jmla.1999.2638

Garber, S., Seigel, G., & Pick, H. (1981). Regulation of vocal intensity in the presence of feedback filtering and amplification. *Journal of Speech and Hearing Research, 24*, 104-108. Retrieved from http://jslhr.highwire.org/cgi/content/abstract/24/1/104

Greenhouse, S.W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95-112. doi: 10.1007/BF02289823

Grieser, D., & Kuhl, P. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 25*, 577-588. doi: 10.1037/0012-1649.25.4.577

Grosjean, F. (1982). *Life with Two Languages.* Cambridge, MA: Harvard University Press.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*, 3-15. doi: 10.1016/0093-934X(89)90048-5

Grosjean, F. (1997). Processing mixed languages: Issues, findings and models. In A. de Groot, & J. Kroll, *Tutorials in Bilingualism: Psycholinguistic Perspectives* (pp. 225-254).Mahwah, NJ: Lawrence Erlbaum.

Gosjean, F., & Soares, G. (1986). Processing mixed languages: Some preliminary findings . In J. Vaid (Ed.), *Language Processing in Bilinguals: Psycholinguistic and Neuropsychological Perspectives* (pp. 145-179). Hillsdale, NJ: Lawrence Erlbaum.

Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory and Cognition, 29*, 503-511. doi: 10.3758/BF03196401

Hagiwara, R. (2006). Vowel production in Winnipeg. *The Canadian Journal of Linguistics, 51*(2/3), 127-41. doi: 10.1353/cjl.2008.0022

Hienz, R., Aleszczyk, C., & May, B. (1996). Vowel discrimination in cats: Acquisition, effects of stimulus level, and performance in noise. *Journal of the Acoustical Society of America, 99*, 3656-3668. doi: 10.1121/1.414980

Heinz, R., Sachs, M., & Sinnot, J. (1981). Discrimination of steady-state vowel by blackbirds and pigeons. *Journal of the Acoustical Society of America, 70*, 699-706. doi: 10.1121/1.386933

Hollett, P. (2006). Investigating St. John's English: Real- and apparent-time perspectives. *The Canadian Journal of Linguistics*, *51*(2/3), 143-160. doi: 10.1353/cjl.2008.0003

Houde, J., & Jordan, M. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, *45*, 295-310. doi: 10.1044/1092-4388(2002.023)

Houde, J., & Jordan, M. (1998). Sensorimotor adaptation in speech production. *Science*, *279*, 1213-1216. doi: 10.1126/science.279.5354.1213

Ioup, G. (1995). Evaluating the need for input enhancement in post-critical period language acquisition. In D. Singleton, & Z. Lengyel (Eds.), *The Age Factor in Second Language Acquisition* (pp. 95-123). Clevedon, UK: Multilingual Matters.

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study in a naturalistic environment. *Studies in Second*

Language Acquisition, 16, 73-98. doi: 10.1017/S0272263100012596

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second Language. *Cognitive Psychology, 21*, 60-99. doi: 10.1016/0010-0285(89)90003-0

Jones, J.A., & Munhall, K..G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *Journal of the Acoustical Society of America*, *108*(3), 1246-1251. doi: 10.1121/1.1288414

Jones, J.A., & Munhall, K.G. (2002). The role of auditory feedback during phonation: Studies of Mandarin tone production. *Journal of Phonetics*, *30*, 303-320. doi: 10.1006/jpho.2001.0160

Kuhl, P. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Attention,* Perception *& Psychophysics , 50*(2), 93-107. doi: 10.3758/BF03212211

Kuhl, P.K. (1986). Reflections on infants' perception and representation of speech. In J.S. Perkell & D.H. Klatt (Eds), *Invariance and Variability in Speech Processes* (pp. 19-30). Hillsdale, NJ: Erlbaum.

Lane, H., & Trainel, B. (1971). Lombard sign and role of hearing in speech. *Journal of Speech and Hearing Research*, *14*, 677-709. Retrieved from http://jslhr.asha.org/cgi/content/abstract/14/4/677

Lenneberg, E. (1967). *Biological Foundations of Language.* New York: John Wiley & Sons, Inc.

Liem, N. (1970). *Vietnamese Pronunciation.* Honolulu, HI: University of Hawaii Press.

Liu, H., Bates, E., & Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Applied Psycholinguistics, 13*, 451-484. doi: 10.1017/S142716400005762

Liu, H., & Larson, C.R. (2007). Effects of perturbation magnitude and voice F0 level on the pitch shift reflex. *Journal of the Acoustical Society of America*, *122*(6), 3671-3677. doi: 10.1121/1.2800254

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*, 251-285. doi: 10.1017/S0272263100009165

MacDonald, E.N., Goldberg, R., & Munhall, K.G. (2010). Compensations in response to real time formant perturbations of different magnitudes. *Journal of the Acoustical Society of America, 127*(2), 1059-1068. doi: 10.1121/1.3278606

MacDonald, E.N, Purcell, D.W., & Munhall, K.G. (2011). Probing the independence of formant control using altered auditory feedback. *Journal of the Acoustical Society of America*, *129*(2), 955-965. doi: 10.1121/1.3531932

Mack, M. (1990). Phonetic transfer in a French-English bilingual child. In P. Nelde (Ed.), *Language Attitudes and Language Conflict* (pp. 187-200). Bonn, Germany: Dummler.

MacKay, I., Flege, J., Piske, T., & Schirru, C. (2001a). Category restructuring during second language (L2) speech acquisition. *Journal of the Acoustical Society of America, 110*(1), 516-528. doi: 10.1121/1.1377287

MacKay, U., Meador, D., & Flege, J. (2001b). The identification of English consonants by native speakers of Italian. *Phonetica, 58*, 103-125. doi: 10.1159/000028490

Max, L., & Onghena, P. (1999). Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research. *Journal of Speech, Language, and Hearing Research, 42*, 261-270. Retrieve from http://jslhr.highwire.org/cgi/content/abstract/42/2/261

McDonald, J. (2000). Grammatically judgments in a second language: Influences of age of acquisition and native language. *Applied Linguistics*, *21*(3), 395-423. doi: 10.1017/S0142716400003064

McLaughin, B. (1977). Second-language learning in children. *Psychological Bulletin, 84*, 438-459. doi: 10.1037/0033-2909.84.3.438

Meador, D., Flege, J., & MacKay, I. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition, 3*, 55-67. doi: 10.1017/S1366728900000134

Mendoz-Denton, N., Hendricks, S., & Kennedy, R. (2001). *Language Samples Project: Canadian English*. Retrieved from http://www.ic.arizona.edu/~lsp/Canadian/ canphon2.html

Miller, J., & Volaitis, L. (1989). Effect of speaking rate on the perceptual structure of phonetic category. *Perception & Psychophysics, 46*, 505-512. doi:

10.3758/BF03208147

Mitsuya, T., MacDonald, E.N., Purcell, D.W., & Munhall, K.G. (2010). *A cross-language study of compensatory response to formant-shifted feedback.* Paper presented at the Second Language Studies – Acquisition, Learning, Education and Technology, Tokyo, Japan. Retrieved from http://www.gavo.t.u tokyo.ac.jp/L2WS2010/papers/L2WS2010_O3-03.pdf

Mitsuya, T., MacDonald, E.N., Purcell., D.W., & Munhall, K.G. (2011). A cross language study of compensation in response to real-time formant perturbation. *Journal of the Acoustical Society of America, 130*(5), 2978-2986. doi: 10.1121/1.3643826

Munhall, K.G., MacDonald, E.N., Byrn, S., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *Journal of the Acoustical Society of America, 125*(1), 384-390. doi: 10.1121/1.3035829

Munro, M., Flege, J., & MacKay, I. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics, 17*, 313-334. doi:10.1017/S0142716400007967

Nasir, S.M., & Ostry, D.J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience*, 11, 1217-1222. doi: 10.1038/nn.2193

Neufeld, G. (1979). Towards a theory of language learning ability. *Language Learning, 29*(2), 227-241. doi: 10.1111/j.1467-1770.1979.tb01066.x

Novoa, L., Fein, D., & Obler, L. (1988). Talent in foreign languages: A case study. In L. Obler, & D. Fein (Eds.), *The Exceptional Brain. Neuropsychology of Talent and Special Abilities* (pp. 294-302). New York: Guilford Press.

O'Shaughnessy, D. (1988). Linear predictive coding. *IEEE Xplore Potentials*, *7*(1), 29 32. doi: 10.1109/45.1890

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research, 5*, 261-283. doi: 10.1007/BF01067377

Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics, 11*, 73-89. doi: 10.1093/applin/11.1.73

Patkowski, M. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning, 30,* 449-472. doi: 10.1111/j.1467-1770.1980.tb00328.x

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. doi: 10.1121/1.1906875

Piske, T., Flege, J., MacKay, I., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica, 59*, 49-71. doi: 10.1159/000056205

Polka, L., & Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English learning and German-learning infants. *Journal of Acoustical Society of America, 100*, 577-592. doi: 10.1121/1.415884

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication, 41*, 221-231. doi: 10.1016/S0167-6393(02)00105-X

Polka, L., & Werker, J. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 421-435. doi: 10.1037/0096-1523.20.2.421

Purcell, D.W., MacDonald, E.N., & Munhall, K.G. (2011). *Cross-vowel use of auditory feedback in English*. Paper presented at the International Seminar on Speech Production (p 166-167), Montreal, QC.

Purcell, D.W., & Munhall, K.G. (2006a). Adaptive control of vowel formant frequency: Evidence from real time formant manipulation. *Journal of the Acoustical Society of America, 120*(2), 966-977. doi: 10.1121/1.2217714

Purcell, D.W., & Munhall, K.G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *Journal of the Acoustical Society of America, 119*(4), 2288-2297. doi: 10.1121/1.2173514

Rosh, E. (1975). Inductive judgements about natural categories. *Journal of Verbal Learning & Verbal Behavior*, 14, 665-681. doi: 10.1016/S0022-5371(75)80055-7

Santry, P. (1997). *The way south Vietnamese pronounce English.* Frankfurt am Main: Wissenchaftliche Buchhandlung T. Hector.

Schneiderman, E., & Desmarais, C. (1988). The talented language learner: Some preliminary findings. *Second Language Research, 4,* 91-109. doi: 10.1177/026765838800400201

Scovel, T. (1988). *A Time to Speak. a Psycholinguistic Inquiry into the Critical Period for Human Speech.* Cambridge, MA: Newbury House Publishers.

Sebastian-Galles, N., & Bosch, L. (2005). Phonology and bilingualism. In J. Kroll, & A. de Groot, *The Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 68-87). New York: Oxford University Press.

Seliger, H. (1978). Implications of a multiple critical period hypothesis for second language learning. In W. Ritchie (Ed.), *Second Language Acquisition Research. Issues and Implications* (pp. 11-19). New York: Academic Press.

Swoboda, P., Kass, J., Morse, P., & Leavitt, L. (1978). Memory factors in infant vowel discrimination of normal and at-risk infants. *Child Development, 49,* 332-339. doi: 10.2307/1128695

Tang, G., & Barlow, J. (2006). Characteristics of the sound systems of monolingual Vietnamese-speaking children with phonological impairment. *Clinical Linguistics &Phonetics*, *20*(6), 423-445. doi: 10.1020/02699200500100910

Tremblay, S., Schiller, D.M., & Ostry, D.J. (2003). Somatosensory basis of speech production. *Nature*, *423*, 866-869. doi: 10.1038/nature01710

Walsh, T., & Diller, K. (1981). Neurolinguistic considerations on the optimum age for second language learning. In K. Diller (Ed.), *Individual Differences and Universals in Language Learning Aptitude* (pp. 3-21). Rowley, MA: Newbury House.

Watt, D. (2006). *Phonetics: Standardised Reading Passages*. Retrieved from http://www.abdn.ac.uk/langling/resources/phonetics.html

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research, 12,* 238-265. doi: 10.1177/026765839601200301

Villacorta, V., Perkell, J., & Guenther, F. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *Journal of the Acoustical Society of America*, *122*(4), 2306-2319. doi: 10.1121/1.2773966

Xu, T., Larson, C., Bauer, J., & Hain, T. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *Journal of the Acoustical Society of America*, *116*(2), 1168-1178. doi: 10.1121/1.1763952

Yeni-Komshian, G., Flege, J., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition*, *3*(2), 131-150. doi: 10.1017/S1366728900000225

Appendices

Appendix A: Ethics approval notice

Research Western

Use of Human Participants - Ethics Approval Notice

Principal Investigator: Dr. David Purcell
Review Number: 17885E
Review Level: Delegated
Approved Local Adult Participants: 65
Approved Local Minor Participants: 0
Protocol Title: Speech production and Auditory Feedback across Languages
Department & Institution: Communication Sciences & Disorders, University of Western Ontario
Sponsor:
Ethics Approval Date: March 16, 2011 **Expiry Date:** August 31, 2012
Documents Reviewed & Approved & Documents Received for Information:

Document Name	Comments	Version Date
UWO Protocol		
Advertisement	Poster	
Advertisement	Flyer	
Letter of Information & Consent		

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the UWO Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The UWO HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Signature

Ethics Officer to Contact for Further Information

___ Janice Sutherland	___ **Elizabeth Wambolt**	✓ **Grace Kelly**

The **University** *of* **Western Ontario**
Office of Research Ethics

www.uwo.ca/research/ethics

Appendix B: Forms and questionnaires for participants

Western
Health Sciences

Speech Production and Auditory Feedback across Languages

David Purcell, Ph.D., Assistant Professor
National Centre for Audiology
School of Communication Sciences and Disorders
University of Western Ontario

LETTER OF INFORMATION

Study Background

You are being invited to participate in a study which investigates how speech production is affected by hearing your own voice, also known as auditory feedback. The study also investigates the usage of auditory feedback in speech production across languages. All measurements will take place in the Electrophysiology Laboratory of the National Centre for Audiology in Elborn College at the University of Western Ontario.

Hearing your own voice is crucial for learning and maintaining accurate speech production. Deafness early in life hampers the development of normal speech in children, and in adults the onset of hearing impairment can cause deterioration in many aspects of speech. This study will attempt to determine the role of the auditory feedback in shaping articulation patterns for vowels where English is a second language.

Questionnaire and Hearing Assessment

This study will include a total of 65 individuals. If you agree to participate in the study, you will take part in a brief questionnaire and a brief assessment of your hearing. This will be followed by the main experiment where you will be prompted to speak into a microphone and to listen to speech through a set of headphones.

The hearing assessment will be a measurement called a pure-tone audiogram which takes about 12 minutes to complete. You will hear tones one at a time through headphones, and you will signal when you detect each tone. The tones will progressively become quieter until you are no longer able to hear them. This procedure is repeated for several different pitches and for each ear.

Western
Health Sciences

Speech Production and Auditory Feedback

During the experiment, you will be seated in front of a video monitor that displays words. You will be asked to say the words on the screen. Your voice will be recorded with a microphone, and played to you through headphones. The microphone and headphones are connected to a computer that can analyze and change the speech sounds. This part of the experiment normally takes less than 45 minutes to complete. You may also be asked to listen to speech sounds and to indicate what you think of those sounds with a button press. This part of the experiment also normally takes less than 30 minutes to complete.

These methods are widely used in laboratories studying hearing and speech. There are no known risks associated with this technology and the equipment meets the highest national and international safety standards.

Benefits to Study Participation

Participation in the study is voluntary. You may refuse to participate, or withdraw from the study at any time. The procedures to be used in this study are designed for research purposes and are not intended to provide you with any direct benefit. It may contribute to our understanding of the role of hearing on speech production and the process of learning languages, which is of benefit to society in the long term. There may be the possibility that the brief hearing assessment could identify a previously unknown hearing impairment. If this were to occur, we will encourage you to seek professional assessment from your family practitioner or audiologist. We may also provide information about obtaining an assessment at the UWO audiology clinic in Elborn College.

All information obtained in this study will be held in strict confidence and participant anonymity will be maintained. Your name will not appear in any publications or presentations of the findings of this study. Your personal and background information will be kept separately from all data. In addition, the data obtained in this study will only be connected via a master list and the Unique ID of each participant. If you would like to receive copies of these publications, please contact Dr. Purcell at the telephone number below.

On the Consent Form that follows, you will be given the opportunity to indicate that you would be interested in receiving invitations to participate in future research studies conducted by Dr. Purcell. This opportunity is unrelated to the present research study.

Western
Health Sciences

If you have any questions or would like additional information about this study, please contact Dr. David Purcell, National Centre for Audiology, School of Communication Sciences and Disorders, University of Western Ontario, London, Ontario, (telephone:

If you have questions regarding the conduct of this study or your rights as a research participant, you may contact the Office of Research Ethics at or via electronic mail at

Compensation

Participants in this study are reimbursed for the time committed to the study and the inconveniences associated with participation in the study at the rate of $5/half-hour or part-thereof.

Signing of Consent Form

If you agree to participate in this study, please sign the consent form. You do not waive any legal rights by signing the consent form. You will be given a copy of this Letter of Information for your records.

Sincerely,

David W. Purcell, Ph.D.
Assistant Professor
National Centre for Audiology
School of Communication Sciences and Disorders
University of Western Ontario

Western
Health Sciences

Speech Production and Auditory Feedback across Languages

David Purcell, Ph.D., Assistant Professor
National Centre for Audiology
School of Communication Sciences and Disorders
University of Western Ontario

CONSENT FORM

I have read the Letter of Information, have had the nature of the study explained to me, and I agree to participate. All questions have been answered to my satisfaction.

☐ Please check the box to the left if we may invite you to participate in future research studies conducted by Dr. David Purcell.

Research Participant (please print):____

Signature:______________________ Date:______________________________

Signature of Person Responsible for Obtaining Signed Consent

Signature: ______________________ Date: ______________________________

Date: ____/____/_____

Background Information

Participant ID: ______________________

Birth year (mm/yyyy): _____/ ______ Age: _________ Sex: Male/Female

Handedness: Right/Left

Vision status: Glasses/Contacts/None

Any known problems with:

i) Hearing: ____

ii) Speech and Language:____

iii) Vision: ____

iv) Other: ____

Date: ____/____/________

Language History and Experience Questionnaire

Participant ID:______________________

(1) What is your country of birth? ____________________________ How many years did you live in Vietnam? __________ At what age did you move to Canada? ___________

(2) What is your 1st Language: _______________
2nd Language: _______________
Father's 1st Language: _________________________
Father's 2nd Languages: ________________________
Mother's 1st Language: _________________________
Mother's 2nd Languages:________________________

List the languages you know in the order:

a) in which you learned them:____
b) from the one you know best to the one you know least:______________________________

Language spoken most frequently at home with your family:__________________________________

(3) What percentage of time are you *currently* exposed to each of the following languages in your daily activities?

Vietnamese _________
English _________
Other _________

(4) What is the main Vietnamese dialect do you speak?

☐ Northern (Ha Noi) ☐ Southern (Saigon) ☐ Central (Hue)

☐ Other (please specify) _______________

(5) Where did you learn (city/province): English _____________
Vietnamese _____________

Experience with English

For each of the following English language skills, please indicate the age at which you first started to acquire the skill, the place in which you learned the skill (e.g. home, school), and rate the fluency with which you can currently perform the skill. (Circle one number per skill).

	starting age	Place	fluency none									very fluent
Understanding			1	2	3	4	5	6	7	8	9	10
Speaking			1	2	3	4	5	6	7	8	9	10
Reading			1	2	3	4	5	6	7	8	9	10
Writing			1	2	3	4	5	6	7	8	9	10

Experience with Vietnamese

For each of the following Vietnamese language skills, please indicate the age at which you first started to acquire the skill, the place in which you learned the skill (e.g. home, school), and rate the fluency with which you can currently perform the skill. (Circle one number per skill).

	starting age	Place	fluency none									very fluent
Understanding			1	2	3	4	5	6	7	8	9	10
Speaking			1	2	3	4	5	6	7	8	9	10
Reading			1	2	3	4	5	6	7	8	9	10
Writing			1	2	3	4	5	6	7	8	9	10

School Experience

Indicate the type of schooling that you received at each grade level by placing an 'x' in the appropriate box.

	K1	K2	1	2	3	4	5	6	7	8	9	10	11	12	U1	U2	U3	U4
Vietnamese school in Vietnam with no English course																		
Vietnamese school in Vietnam with an English course																		
English school with no Vietnamese course																		
English school with a Vietnamese course																		

Music History

Participant ID: ______________________________ Date:_______________

Have you had vocal (singing) lessons: YES NO

If yes, what type of training did you receive:

____

If yes, what was your highest completion of grade/level/number of years:

____

Have you had instrument lessons: YES NO

If yes, what was your instrument(s):

____

If yes, what was your highest completion of grade/level/number of years:

____

Appendix C: English reading passages

1) **The grandfather passage (Watt, 2006)**
You wished to know all about my grandfather. Well, he is nearly ninety-three years old; he dresses himself in an ancient black frock coat, usually minus several buttons; yet he still thinks as swiftly as ever. A long, flowing beard clings to his chin, giving those who observe him a pronounced feeling of the utmost respect. When he speaks, his voice is just a bit cracked and quivers a trifle. Twice each day he plays skilfully and with zest upon our small organ. Except in the winter when the ooze or snow or ice prevents, he slowly takes a short walk in the open air each day. We have often urged him to walk more and smoke less, but he always answers, “Banana oil!” Grandfather likes to be modern in his language.

2) **The North Wind and the Sun (Watt, 2006)**
The North Wind and the Sun were disputing which was the stronger, when a traveller came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveller take his cloak off should be considered stronger than the other. Then the North Wind blew as hard as he could, but the more he blew the more closely did the traveller fold his cloak around him, and at last the North Wind gave up the attempt. Then the Sun shone out warmly, and immediately the traveller took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two.

3) **The Rainbow Passage (Watt, 2006)**
When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colours. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow. Throughout the centuries people have explained the rainbow in various ways. Some have accepted it as a miracle without physical explanation. To the Hebrews it was a token that there would be no more universal floods. The Greeks used to imagine that it was a sign from the gods to foretell war or heavy rain. The Norsemen considered the rainbow as a bridge over which the gods passed from earth to their home in the sky. Others have tried to explain the phenomenon physically. Aristotle thought that the rainbow was caused by reflection of the sun's rays by the rain. Since then physicists have found that it is not reflection, but refraction by the raindrops which causes the rainbows. Many complicated ideas about the rainbow have been formed. The difference in the rainbow depends considerably upon the size of the drops, and the width of the colored band increases as the size of the drops increases. The actual primary rainbow observed is said to be the effect of super-imposition of a number of bows. If the red of the second bow falls upon the green of the first, the result is to give a bow with an abnormally wide yellow band, since red and green light when

mixed form yellow. This is a very common type of bow, one showing mainly red and yellow, with little or no green or blue.

4) Arthur the Rat (Watt, 2006)

Once upon a time there was a rat who couldn't make up his mind. Whenever the other rats asked him if he would like to come out hunting with them, he would answer in a hoarse voice, "I don't know." And when they said, "Would you rather stay inside?" he wouldn't say yes, or no either. He'd always shirk making a choice.

One fine day his aunt Josephine said to him, "Now look here! No one will ever care for you if you carry on like this. You have no more mind of your own than a greasy old blade of grass!" The young rat coughed and looked wise, as usual, but said nothing

"Don't you think so?" said his aunt stamping with her foot, for she couldn't bear to see the young rat so coldblooded.

"I don't know," was all he ever answered, and then he'd walk off to think for an hour or more, whether he would stay in his hole in the ground or go out into the loft.

One night the rats heard a loud noise in the loft. It was a very dreary old place. The roof let the rain come washing in, the beams and rafters had all rotted through, so that the whole thing was quite unsafe.

At last one of the joists gave way, and the beams fell with one edge on the floor. The walls shook, and the cupola fell off, and all the rats' hair stood on end with fear and horror.

5) English text passage from OT-APST (Cooke, n.d.):

Movie and art give pleasure to all age groups. People
from all over the world dance to the feel of
the music. You can learn about the music and art of a culture from
the people and by reading books. Children sing together from a young age
and move to the rhythm. Parents will sit
and watch the children play. The study of music improves school
success. The school band has students with many skills
and much talent. Listen to the band. Clap your hands and tap
your feet. The crowd will cheer in support. Art galleries
and enjoy the beauty of colour. Music is the food of the soul
and art is the wine.

Appendix D: Vietnamese reading passages

1) The North Wind and the Sun (Bradlow, 2010)

Gió Bắc và Mặt trời có một cuộc thi xem thử ai là người có thể làm cho nhà du hành cởi áo khóac ra. Gió Bắc hùng mạnh ra sức thổi, nhưng rốt cuộc cũng chỉ làm cho nhà du hành càng ôm chiếc áo sát vào người mà thôi. Ngược lại, Mặt trời nhẹ nhàng chiếu những tia sáng của mình vào nhà du hành. Được sưởi ấm, nhà du hành cảm thấy không còn lý do gì để mặc áo khóac nữa, và cởi nó ra.

2) Vietnamese text passage from OT-APST (Cooke, n.d.)

Âm nhạc và nghệ thuật đem sảng khoái đến mọi lứa tuổi. Người ta
trên khắp thế giới nhảy múa theo cảm nhận của
âm nhạc . Bạn có thể học nhạc và nghệ thuật của một nền văn hóa từ
người ta và bằng cách đọc sách. Trẻ con hát chung với nhau từ tuổi nhỏ
và di chuyển theo điệu nhạc. Cha mẹ ngồi
và xem trẻ con chơi. Môn học âm nhạc làm tăng sự thành công của
trường. Ban nhạc của trường có nhiều học sinh với nhiều tài năng
và rất tài giỏi. Lắng nghe ban nhạc. Vỗ tay và nhịp
chân của bạn. Đám đông sẽ reo lên ủng hộ. Phòng triển lãm nghệ thuật
là nơi mọi người có thể thư thản
và thưởng thức vẻ đẹp của màu sắc. Âm nhạc là thức ăn của linh hồn
và nghệ thuật là rượu.

Curriculum Vitae

Name	Linh Le Truc Nguyen
Post-secondary Education and Degrees	The University of Western Ontario London, Ontario 2010-2012, MSc
	The University of Toronto Mississauga, Ontario 2006-2010, HBSc
Honours and Awards	Province of Ontario Graduate Scholarship 2011-2012
	First Place, Oral Presentation Western Graduate Research Forum The University of Western Ontario 2012
	Health and Rehabilitation Sciences Travel Award The University of Western Ontario 2012
	Faculty of Health Sciences Travel Award The University of Western Ontario 2011
Related Work Experience	Undergraduate Teaching Assistant Faculty of Health Sciences The University of Western Ontario 2011
Forum and Conference Experiences	HRS Graduate Research Forum Committee The University of Western Ontario 2011-2012
Other Related Experience	Student Services Committee The University of Western Ontario 2011-2012, 2012-2013
	School of Graduate Studies Councilor The University of Western Ontario 2011-2012

Student Field Mentor (Hearing Science)
The University of Western Ontario
2011-2012

HRS Graduate Student Society (VP Finance)
The University of Western Ontario
2011-2012

HRS Community Involvement Committee
The University of Western Ontario
2010-2011

Conference (International) Presentations

Nguyen, L.L.T.*, & Purcell, D.W. (212, May). *Formant compensation responses to altered auditory feedback in English and Vietnamese talkers*. Oral presentation at the Listening Talker Workshop, Edinburgh, UK.

Nguyen, L.L.T.*, & Purcell, D.W. (2012, Mar). *Speech compensation during altered auditory feedback in Canadian and Vietnamese vowels.* Oral presentation at the Western Interdisciplinary Student Symposium on Language Research, London, ON.

Nguyen, L.L.T.*, & Purcell, D.W. (2011, Dec). *A comparison of speech production and perception in English and Vietnamese speaker.* Poster presentation at the Bilingual Workshops in Theoretical Linguistics, Hamilton, ON.

Ben-David, B,M., & **Nguyen, L.L.T.*** (2009, Nov). *Stroop effects in patients with traumatic brain injury: Selective attention or speed of processing? A review*. Poster presentation at the Object, Perception, Attention and Memory Conference, Boston, MA.

Conference (Institutional) Presentations

Nguyen, L.L.T.*, & Purcell, D.W. (2012, Mar*). A comparison of speech production and perception in English and Vietnamese speakers*. Poster presentation at the Faculty of Heath Sciences Research Day, London, ON.

Nguyen, L.L.T.*, & Purcell, D.W. (2012, Mar). *Cross-language study in speech compensation responses to altered auditory feedback*. Oral presentation at the Western Graduate Research Forum, London, ON.

Nguyen, L.L.T.*, & Purcell, D.W. (2012, Mar*). A comparison of speech production and perception in English and Vietnamese speakers*. Poster presentation at the Western Graduate Research Forum, London, ON.

Nguyen, L.L.T.*, & Purcell, D.W. (2012, Feb). *A comparison of speech production and perception in English and Vietnamese speakers.* Poster presentation at the Health and Rehabilitation Sciences Graduate Research Forum, London, ON.

Public Presentations

Nguyen, L.L.T.*, & Purcell, D.W. (2012, Mar). *Relationship between vowel categorization and speech compensation from altered auditory feedback in*

Canadian and Vietnamese talkers. Oral presentation at the Research in Hearing Science Seminar Series, London, ON.

Nguyen, L.L.T.*, & Purcell, D.W. (2011, Nov). *Preliminary look into altered auditory feedback in English and Vietnamese Speakers.* Oral presentation at the Research in Hearing Science Seminar Series, London, ON.

Nguyen, L.L.T.*, & Purcell, D.W. (2011, Mar). *Acquisitions and interactions between first and second-languages.* Oral presentation at the Research in Hearing Science Seminar Series, London, ON.

Nguyen, L.L.T.*, & Purcell, D.W. (2011, Nov). *Comparison of English and Vietnamese vowels*. Oral presentation at the Research in Hearing Science Seminar Series, London ON.

Peer-Reviewed Journal Publication

Ben-David, B.M., **Nguyen, L.L.T**., & van Lieshout, P.H.H.M. (2011). Stroop effects in persons with traumatic brain injury: Selective attention, speed of processing or color-naming? A meta-analysis. *Journal of the International Neuropsychological Society*, 17, 354-363.

Short Paper and Abstract Publications

Nguyen, L.L.T. & Purcell, D.W. (2012, May). Formant compensation responses to altered auditory feedback in English and Vietnamese talkers. *Listening Talker Workshop*, Edinburgh, Scotland. Retrieved from http://listening-talker.org/workshop/abstracts/oral/F3.pdf

Nguyen, L.L.T. & Purcell, D.W. (2012, Feb). A comparison of speech production and perception in English and Vietnamese speakers. *Health and Rehabilitation Sciences Graduate Research Forum* (p. 33), London, Ontario.

Ben-David, B. M**.** & **Nguyen, L.L.T.** (Nov 2009). Stroop effects in patients with traumatic brain injury: Selective attention or speed of processing? A review. *Proceedings of the 17th Annual Meeting of Object Perception Attention and Memory association- OPAM* (p. 22), Boston, MA.